



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 116716

TO: Vanessa L Ford
Location: REM/3B25/3C18
Art Unit: 1645
Friday, March 12, 2004

Case Serial Number: 09/543407

From: Barb O'Bryen
Location: Biotech-Chem Library
Remsen E01A69
Phone: 571-272-2518 *BOB*

barbara.obryen@uspto.gov

Search Notes

RUSH

THIS PAGE BLANK (USPTO)

116716

From: Chan, Christina
Sent: Thursday, March 11, 2004 12:51 PM
To: Ford, Vanessa; STIC-Biotech/ChemLib
Subject: RE: In re: 09543407 sequence search

Please rush. Thanks Chris

Chris Chan

TC 1600 New Hire Training Coordinator and SPE 1644
(571)-272-0841
Remsen, 3E89

-----Original Message-----

From: Ford, Vanessa
Sent: Thursday, March 11, 2004 11:40 AM
T : Chan, Christina
Subject: In re: 09543407 sequence search

Please search SEQ ID NO: 12, 14, 16, 18, 20, 22, 24, 26, 28 and 30. Please include interference searches. Please rush.

Vanessa L. Ford
Biotechnology Patent Examiner
Office: REM 3B25
Mailbox: REM 3C18
Phone: 571.272.0857
Art Unit: 1645

RECEIVED
MAR 11 2004
STIC-BIOTECH/ChemLib
(STIC)

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: _____
Date Completed: _____
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

THIS PAGE BLANK (USPTO)

CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell) the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

SQ Sequence 151 AA;
 Query Match
 Best Local Similarity 100.0%; Score 779; DB 3; Length 151;
 Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MLLKVAAPFAIVSGSALAGVPOWGCGGNHNGGNSGPDSTLSIYOGSANAALALQ 60
 DB 1 MLLKVAAPFAIVSGSALAGVPOWGCGGNHNGGNSGPDSTLSIYOGSANAALALQ 60
 QY 61 SDARSETTTTQSGYGADVGAGADNSTLTETONGFRNNAATIDONNANNSDITVGYG 120
 DB 61 SDARSETTTTQSGYGADVGAGADNSTLTETONGFRNNAATIDONNANNSDITVGYG 120
 QY 121 NNAALVNVYDQVTRVYVTHEMAHANNATANOY 151
 DB 121 NNAALVNVYDQVTRVYVTHEMAHANNATANOY 151

RESULT 2
 AAB36347
 ID AAB36347 standard; protein; 151 AA.
 AC AAB36347;
 XX 26-FEB-2001 (first entry)
 DE Agfa::PT#2 amino acid sequence SEQ ID NO:14.
 XX
 KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 PN WO200060102-A2.
 PD 12-OCT-2000.
 PF 05-APR-2000; 2000WO-CA000356.
 PR 05-APR-1999; 99US-0127888P.
 XX (UYVI-) UNIV VICTORIA.
 PA White AP, Doran JL, Collison SK, Kay WW;
 DR WPI; 2000-672631/65.
 DR N-PSDB; AAC64623.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 136; 139pp; English.

XX The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/7MF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell); the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

SQ Sequence 151 AA;
 Query Match
 Best Local Similarity 89.3%; Score 696; DB 3; Length 151;
 Matches 141; Conservative 0; Mismatches 0; Indels 20; Gaps 2;
 QY 1 MLLKVAAPFAIVSGSALAGVPOWGCGGNHNGGNSGPDSTLSIYOGSANAALALQ 60
 DB 1 MLLKVAAPFAIVSGSALAGVPOWGCGGNHNGGNSGPDSTLSIYOGSANAALALQ 60
 QY 61 SDARSETTTTQSGYGADVGAGADNSTLTETONGFRNNAATIDONNANNSDITVGYG 120
 DB 61 SDARSETTTTQSGYGADVGAGADNSTLTETONGFRNNAATIDONNANNSDITVGYG 120
 QY 121 NNAALVNVYDQVTRVYVTHEMAHANNATANOY 151
 DB 118 YDQVTRVYVTHEMAHANNATANOY 151

RESULT 3
 AAR74625
 ID AAR74625 standard; protein; 151 AA.
 AC AAR74625;
 XX 25-MAR-2003 (revised)
 DT 26-JUN-1995 (first entry)
 DE Agfa sequence.
 XX
 KW Salmonella; Agfa; vaccine.
 KW Salmonella.
 OS
 PN WO9425598-A2.
 PD 10-NOV-1994.
 PF 26-APR-1994; 94WO-IB000207.
 PR 26-APR-1993; 93US-00054452.
 XX (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
 PA (KING/) KING J.

PI Kay WW, Collinson SK, Clouthier SC, Doran JL;
 DR WPI; 1994-358275/44.
 DR N-PSDB; AAC87467.
 XX
 PT Eliciting an immune response to *Salmonella* - using attenuated *Salmonella*
 PT strains; vector constructs, or compens. contg. fimbrial type proteins.
 XX
 PS Disclosure; Fig 7B; 95pp; English.
 CC The *Salmonella* Agfa protein and DNA are used in vaccine and genetic
 CC immunization compositions, respectively, to elicit an immune response to
 CC *Salmonella* in animals (e.g. food producing animals) and humans. (Updated
 CC on 25-MAR-2003 to correct PW field.)
 XX
 SQ Sequence 151 AA;
 Query Match 89.2%; Score 695; DB 2; Length 151;
 Best Local Similarity 90.1%; Pred. No. 2.2e-58;
 Matches 136; Conservative 4; Mismatches 11; Indels 0; Gaps 0;
 QY 1 MLLKVAFAFAIVSGSALAGVVPWGGGNNHGGNSGDPSTLSIYOGSANAALAQ 60
 DB 1 MLLKVAFAFAIVSGSALAGVVPWGGGNNHGGNSGDPSTLSIYOGSANAALAQ 60
 QY 61 SPARKSETTITGSGNGADVGGADNSTIELTONGFRNNATIDOMNANSDITVGOYGG 120
 DB 61 SPARKSETTITGSGNGADVGGADNSTIELTONGFRNNATIDOMNANSDITVGOYGG 120
 QY 121 NNAALVNTDQVTRVVTHEMAHANNATANQY 151
 DB 121 NNAALVNTDQVTRVVTHEMAHANNATANQY 151
 RESULT 4
 ID AAB36341 standard; protein; 151 AA.
 XX
 AC AAB36341;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE *Salmonella enteritidis* Agfa amino acid sequence SEQ ID NO:5.
 XX
 KW *Salmonella*; agfa; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS *Salmonella enteritidis*.
 XX
 PN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CN000356.
 XX
 PR 05-APR-1999; 99US-012788BP.
 XX
 PA (UUVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collinson SK, Kay WW;
 XX
 DR WPI; 2000-672631/65.
 DR N-PSDB; AAC64617.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 135; 139pp; English.
 CC The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:

CC (1) use of thin aggregative fimbriae (SEF17/TA) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa; Caga and Agfa-homologue fimbria subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *B. coli* or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;
 Query Match 89.2%; Score 695; DB 3; Length 151;
 Best Local Similarity 90.1%; Pred. No. 2.2e-58;
 Matches 136; Conservative 4; Mismatches 11; Indels 0; Gaps 0;
 QY 1 MLLKVAFAFAIVSGSALAGVVPWGGGNNHGGNSGDPSTLSIYOGSANAALAQ 60
 DB 1 MLLKVAFAFAIVSGSALAGVVPWGGGNNHGGNSGDPSTLSIYOGSANAALAQ 60
 QY 61 SPARKSETTITGSGNGADVGGADNSTIELTONGFRNNATIDOMNANSDITVGOYGG 120
 DB 61 SPARKSETTITGSGNGADVGGADNSTIELTONGFRNNATIDOMNANSDITVGOYGG 120
 QY 121 NNAALVNTDQVTRVVTHEMAHANNATANQY 151
 DB 121 NNAALVNTDQVTRVVTHEMAHANNATANQY 151
 RESULT 5
 ID AAW23570 standard; protein; 151 AA.
 XX
 AC AAW23570;
 XX
 DT 25-MAR-2003 (revised)
 DT 29-SEP-1997 (first entry)
 XX
 DE *Salmonella enteritidis* 27655-3b agfa.
 XX
 KW Enteropathogenic bacteria; enterobacteria; *S. enteritidis*; antibody.
 KW *Salmonella enteritidis*.
 XX
 OS *Salmonella enteritidis*.
 XX
 PN US5635617-A.
 XX
 PD 03-JUN-1997.
 XX
 PF 26-APR-1994; 94US-00233788.
 XX
 PR 26-APR-1993; 93US-00054452.
 XX
 PA (UUVI-) UNIV VICTORIA INNOVATION & DEV CORP.
 XX
 PI Collinson SK, Kay WW, Doran JL;

```

XX PT WPI; 1997-309886/28.
DR DR N-PSDB; AAT74142.
XX XX
PT Isolated Salmonella gene agfa - used for diagnosis of Salmonella or
XX enteropathogenic bacteria of the Enterobacteriaceae family.
PS Example 2; Fig 7, 85pp; English.
XX
CC The present sequence represents agfa encoded by the full agfa gene
CC derived from Salmonella enteritidis 27655-3b. The nucleic acid can be
CC used to provide diagnostic assays for Salmonella and/or enteropathogenic
CC bacteria of the family Enterobacteriaceae. It can also be used to provide
CC proteins and antibodies which can be used for assays. The nucleic acid
CC sequence can be used to provide probes or primers which can specifically
CC hybridise to nucleic acid molecules from greater than 99% of Salmonella
CC strains that are pathogenic to warm-blooded animals relative to nucleic
CC acid molecules from virtually all other microbial organisms. (Updated on
CC 25-MAR-2003 to correct PF field.)
XX
SQ Sequence 151 AA;
Query Match 88.6%; Score 690; DB 2; Length 151;
Best Local Similarity 89.4%; Pred. No. 6.5e-58;
Matches 135; Conservative 4; Mismatches 12; Indels 0; Gaps 0;
QY 1 MLLKVAFAFALIVSGSALAGVVPQGGGNNHNGGNSGDPSTLSTIYQGSANALALQ 60
DB 1 MLLKVAFAFALIVSGSALAGVVPQGGGNNHNGGNSGDPSTLSTIYQGSANALALQ 60
QY 61 SDARKSETTITGSGYGNAGADVQAGADNSTIELTONGFRNNATIDQWNAKNSDITVGYG 120
DB 61 SDARKSETTITGSGYGNAGADVQAGADNSTIELTONGFRNNATIDQWNAKNSDITVGYG 120
QY 121 NNAALVNTDQLVTRVTHEMAHANNTANQY 151
DB 121 NNPALVNOTASDSSVWVRQVFGNNATNAY 151
RESULT 6
AAB36352
ID AAB36352 standard; protein; 151 AA.
XX
AC AAB36352;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa::PT3#7 amino acid sequence SEQ ID NO:24.
XX
KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
KW vaccine; immune response; immunogen.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UUVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX WPI; 2000-672631/65.
XX DR N-PSDB; AAC64628.
XX
PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
which encodes foreign epitope or antigen, expresses recombinant Agfa

```

```

PT protein useful for eliciting immune response in animal.
XX
BS Disclosure; Page 138, 139pp; English.
XX
CC The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/7AP) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, GsaA and Agfa-homologue fimbria subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (II) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 151 AA;
Query Match 84.1%; Score 655; DB 3; Length 151;
Best Local Similarity 79.8%; Pred. No. 1.4e-54;
Matches 134; Conservative 0; Mismatches 0; Indels 34; Gaps 2;
QY 1 MLLKVAFAFALIVSGSALAGVVPQGGGNNHNGGNSGDPSTLSTIYQGSANALALQ 60
DB 1 MLLKVAFAFALIVSGSALAGVVPQGGGNNHNGGNSGDPSTLSTIYQGSANALALQ 60
QY 61 SDARKSETTITGSGYGNAGADVQAGADNSTIELTONGFRNNATIDQWNAKNSDITVGYG 120
DB 61 SDARKSETTITGSGYGNAGADVQAGADNSTIELTONGFRNNATIDQWNAKNSDITVGYG 120
QY 121 NNAALVNTDQLVTRVTHEMAHA-----NNAATNAY 151
DB 110 -----NNDQLVTRVTHEMAHANQTSADSSVWVRQVFGNNATNAY 151
RESULT 7
AAB36353
ID AAB36353 standard; protein; 151 AA.
XX
AC AAB36353;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa::PT3#8 amino acid sequence SEQ ID NO:26.
XX
KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
KW vaccine; immune response; immunogen.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.

```

PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay WM;
 XX WPI, 2000-672631/65.
 DR N-PSDB; AAC64629.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 138, 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/RAF) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 XX Sequence 151 AA:
 SQ
 Query Match 78.8%; Score 614; DB 3; Length 151;
 Best Local Similarity 80.8%; Pred. No. 1.2e-50;
 Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;
 QY 1 MKLLKVAAPFAIIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
 DB 1 MKLLKVAAPFAIIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALYDQ 60
 QY 61 SDAKSETTTTOSGNGAGADVGGADNSTELTQNGFRNNATTDQWAKNSDITVGOYGG 120
 DB 61 LVTRVVTHERAHAGYGGADVGGADNSTELTQNGFRNNATTDQWAKNSDITVGOYGG 120
 QY 121 NNAALVNYDQLVTRVVTHERAHANATATANOY 151
 DB 121 NNAALVNYQTASDSVVRQYGFGNATATANOY 151
 RESULT 8
 AAB36349
 ID AAB36349 standard; protein; 151 AA.
 AC AAB36349;
 XX
 XX 26-FEB-2001 (first entry)
 XX
 XX Agfa::PT3#4 amino acid sequence SEQ ID NO:18.
 DE
 XX
 XX *Salmonella*; agfa; chromosomal gene replacement; fimbrial; epitope;
 KW vaccine; immune response; immunogen.

XX
 OS *Salmonella enteritidis*.
 OS *Escherichia coli*.
 OS Synthetic.
 XX
 PN W0200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000MO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay WM;
 XX WPI, 2000-672631/65.
 DR N-PSDB; AAC64629.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 136, 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/RAF) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 XX Sequence 151 AA:
 SQ
 Query Match 78.7%; Score 613; DB 3; Length 151;
 Best Local Similarity 80.8%; Pred. No. 1.5e-50;
 Matches 122; Conservative 7; Mismatches 22; Indels 0; Gaps 0;
 QY 1 MKLLKVAAPFAIIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
 DB 1 MKLLKVAAPFAIIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
 QY 61 SDAKSETTTTOSGNGAGADVGGADNSTELTQNGFRNNATTDQWAKNSDITVGOYGG 120
 DB 61 SDAKSETTTTOSGNGAGADVGGADNSTELTQNGFRNNATTDQWAKNSDITVGOYGG 120
 QY 121 NNAALVNYDQLVTRVVTHERAHANATATANOY 151
 DB 121 NNAALVNYQTASDSVVRQYGFGNATATANOY 151
 RESULT 9

ID	AAB36350	standard; protein; 151 AA.
XX	AC	AAB36350;
XX	DT	26-FEB-2001 (first entry)
XX	DE	Agfa::PT3#5 amino acid sequence SEQ ID NO:20.
XX	KW	Salmonella; agfA; chromosomal gene replacement; fimbria; epitope;
XX	XX	vaccine; immune response; immunogen.
OS	Salmonella enteritidis.	
OS	Escherichia coli.	
XX	Synthetic.	
PN	WO20060102-A2.	
XX	PD	12-OCT-2000.
PF	05-APR-2000; 2000MO-CA000356.	
PR	05-APR-1999; 99US-0127888P.	
PA	(UYVI-) UNIV VICTORIA.	
F1	White AP, Doran JL, Collison SK, Kay WW;	
DR	WPI; 2000-672631/65.	
DR	N-PDSB; AAC64626.	
PT	Recombinant agfA gene having a segment replaced by a foreign DNA sequence	
PT	which encodes foreign epitope or antigen, expresses recombinant AgfA	
PS	protein useful for eliciting immune response in animal.	
PS	Disclosure; Page 137; 139pp; English.	
CC	The present invention describes a recombinant agfA gene (1) where a	
CC	segment of the gene has been replaced by a segment of a foreign DNA	
CC	sequence which encodes a foreign epitope or antigen. Also described are:	
CC	(1) use of thin aggregative fimbriae (SEF17/WAF) nucleation depended	
CC	assembly system of strains of Salmonella, Escherichia coli and	
CC	Enterobacteriaceae for the production of fimbriae comprising recombinant	
CC	AgfA, GsgA and AgfA-homologue fimbria subunits, respectively; (2)	
CC	directing recombination of a recombinant gene into the chromosome of the	
CC	homologous species; (3) directing recombination of a recombinant gene	
CC	back into the chromosome of the homologous species, replacing the native	
CC	copy of that gene; and (4) eliciting an immune response in an animal,	
CC	comprising separating an amino acid polymer comprising a recombinant AgfA	
CC	protein containing a replacement segment or segments of foreign amino	
CC	acid sequence or sequences grown on a Salmonella, E. coli or	
CC	enterobacteriaceae host cell, from the host cell and introducing the	
CC	polymer into the animal in conjunction with a carrier or diluent. (1) is	
CC	useful for the expression of recombinant AgfA protein which is useful for	
CC	eliciting an immune response in an animal. In a fibrial presentation	
CC	system the heterologous antigens are presented in high numbers (up to	
CC	500,000 copies/cell), the hybrid fimbria protein possesses both the	
CC	immunogenicity and adhesion properties relevant for an efficient live	
CC	immogens, which may be important for directing an immune response	
CC	against the inserted epitope, and hybrid fimbriae are easy and	
CC	inexpensive to purify in large amount. The present sequence is given in	
CC	the exemplification of the present invention	
XX	Sequence 151 AA;	
XX	Seq	
QY	Query Match	78.4%; Score 611; DB 3; Length 151;
Jb	Best Local Similarity	80.8%; Pred. No. 2.3e-50;
	Matches 122; Conservative 7; Mismatches 22; Indels 0; Gaps 0;	
	1 MKLIRKVAFAAIIVVSGSALAGVVPOWGCGGHHNGGSSGGPDLTSLTYOYSANAAATATQ 60	
	1 MKLIRKVAFAAIIVVSGSALAGVVPOWGCGGHHNGGSSGGPDLTSLTYOYSANAAATATQ 60	

QY 61. SPARSETTITGSGGNCADVGGGADNSTIELTONGERNATTIDOWNAKNSDITVGGYGG 120
 Db 61. SPARKKIDLVTRVVTTHENAHAGGADNSTIELTONGERNATTIDOWNAKNSDITVGGYGG 120
 QY 121. NNPAALVNYDOLVTRVVTTHENAHANNATANY 151
 Db 121. NNPAALVNYDOLVTRVVTTHENAHANNATANY 151
 RESULT 10
 ID AAB36354 standard; protein; 151 AA.
 AC AAB36354;
 DT 26-FEB-2001 (first entry)
 DE Agfa::PT3#9 amino acid sequence SEQ ID NO:28.
 KM Salmoneila; agfa; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 OS Salmoneila enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 PN W0200060102-A2.
 PD 12-OCT-2000.
 PF 05-APR-2000; 2000MO-CA000356.
 PR 05-APR-1999; 99US-0127888P.
 PA (UYVI-) UNIV VICTORIA.
 PI White AP, Doran JL, Collison SK, Kay MW;
 DR WPI; 2000-672631/65.
 DX N-PSDB; AAC64630.
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 PS Disclosure; Page 138; 139pp; English.
 CC The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/RAF) nucleation depended
 CC assembly system of strains of Salmoneila, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmoneila, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in

CC the exemplification of the present invention
XX Sequence 151 AA;

Query Match 77.5%; Score 605; DB 3; Length 151;

Best Local Similarity 80.1%; Pred. No. 8.5e-50; Mismatches 25; Indels 0; Gaps 0;

Matches 121; Conservative 5; Mismatches 25; Indels 0; Gaps 0;

QY 1 MKLLKVAAPFAIVVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANPALALQ 60
DB 1 MKLLKVAAPFAIVVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANPALALQ 60

QY 61 SDARKSETTITGSGYNGADVGAGADNSTIELTQNGFRNNAITDQWNAKNSDITVGQYGG 120
DB 61 SDARKSETTITGSGYNGADVGAGADNSTIELTQNGFRNNAITDQWNAKNSDITVGQYGG 120

QY 121 NNAALVNVYDQLVTRVYTHEMAHANNATANOY 151
DB 121 NNAALVNVYDQLVTRVYTHEMAHANNATANOY 151

RESULT 11

AAB36351 standard; protein; 151 AA.

AC AAB36351;

DT 26-FEB-2001 (first entry)

DE Agfa::PT3#6 amino acid sequence SEQ ID NO:22.

KM Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
KM vaccine; immune response; immunogen.

OS Salmonella enteritidis.

OS Escherichia coli.

OS Synthetic.

PN WO200060102-A2.

PD 12-OCT-2000.

PF 05-APR-2000; 2000WO-CA000356.

PR 05-APR-1999; 99US-0127888P.

PA (UYVI-) UNIV VICTORIA.

PI White AP, Doran JL, Collison SK, Kay MW;

XX WPI; 2000-672631/65.

DR N-PSDB; AAC64627.

XX

PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence

PT which encodes foreign epitope or antigen, expresses recombinant Agfa

PT protein useful for eliciting immune response in animal.

XX

PS Disclosure; Page 137; 139pp; English.

XX

CC The present invention describes a recombinant agfa gene (1) where a

CC segment of the gene has been replaced by a segment of a foreign DNA

CC sequence which encodes a foreign epitope or antigen. Also described are:

CC (1) use of thin aggregative fimbriae (SFP17/7AF) nucleation depended

CC assembly system of strains of Salmonella, Escherichia coli and

CC Enterobacteriaceae for the production of fimbriae comprising recombinant

CC Agfa, C99A and Agfa-homologue fimbria subunits, respectively; (2)

CC directing recombination of a recombinant gene into the chromosome of the

CC homologous species; (3) directing recombination of a recombinant gene

CC back into the chromosome of the homologous species, replacing the native

CC copy of that gene; and (4) eliciting an immune response in an animal,

CC comprising separating an amino acid polymer comprising a recombinant Agfa

CC protein containing a replacement segment or segments of foreign amino

CC acid sequence or sequences grown on a Salmonella, E. coli or

CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention

SQ Sequence 151 AA;

Query Match 77.5%; Score 604; DB 3; Length 151;

Best Local Similarity 80.8%; Pred. No. 1.1e-49; Mismatches 23; Indels 0; Gaps 0;

Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAAPFAIVVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANPALALQ 60
DB 1 MKLLKVAAPFAIVVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANPALALQ 60

QY 61 SDARKSETTITGSGYNGADVGAGADNSTIELTQNGFRNNAITDQWNAKNSDITVGQYGG 120
DB 61 SDARKSETTITGSGYNGADVGAGADNSTIELTQNGFRNNAITDQWNAKNSDITVGQYGG 120

QY 121 NNAALVNVYDQLVTRVYTHEMAHANNATANOY 151
DB 121 NNAALVNVYDQLVTRVYTHEMAHANNATANOY 151

RESULT 12

AAB36355 standard; protein; 151 AA.

AC AAB36355;

DT 26-FEB-2001 (first entry)

DE Agfa::PT3#10 amino acid sequence SEQ ID NO:30.

KM Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
KM vaccine; immune response; immunogen.

OS Salmonella enteritidis.

OS Escherichia coli.

OS Synthetic.

PN WO200060102-A2.

PD 12-OCT-2000.

PF 05-APR-2000; 2000WO-CA000356.

PR 05-APR-1999; 99US-0127888P.

PA (UYVI-) UNIV VICTORIA.

PI White AP, Doran JL, Collison SK, Kay MW;

XX WPI; 2000-672631/65.

DR N-PSDB; AAC64631.

XX

PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence

PT which encodes foreign epitope or antigen, expresses recombinant Agfa

PT protein useful for eliciting immune response in animal.

XX

PS Disclosure; Page 139; 139pp; English.

XX

CC The present invention describes a recombinant agfa gene (1) where a

CC segment of the gene has been replaced by a segment of a foreign DNA

CC sequence which encodes a foreign epitope or antigen. Also described are:

(1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

Sequence 151 AA;

Query Match 77.4%; Score 603; DB 3; Length 151;
 Best Local Similarity 80.8%; Pred. No. 1.3e-49;

Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;

QY 1 MLLKVAFAFAIVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQGSANNAALQ 60
 DB 1 MLLKVAFAFAIVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQGSANNAALQ 60
 QY 61 SDARKSETTITQSGYGADVGAGADNSTIELTONGFRNNATIDQWNAKNSDITVGYGG 120
 DB 61 SDARKSETTITQSGYGADVGAGADNSTIELTONGFRNNATIDQWNAKNSDITVGYGG 120
 QY 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151
 DB 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151

RESULT 13
 AAB36348
 ID AAB36348 standard; protein; 151 AA.
 XX
 AC AAB36348;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Agfa:PT3#3 amino acid sequence SEQ ID NO:16.
 XX
 KW *Salmonella*, agfa; chromosomal gene replacement; fimbrial; epitope;
 XX vaccine; immune response; immunogen.
 OS *Salmonella enteritidis*.
 OS *Escherichia coli*.
 OS Synthetic.
 XX
 PN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UUVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay WW,
 XX WPI; 2000-672631/65.
 DR

DR N-PSDB; AAC64624.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 136; 139pp; English.

The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

Sequence 151 AA;

Query Match 74.5%; Score 580; DB 3; Length 151;
 Best Local Similarity 80.1%; Pred. No. 2.1e-47;

Matches 121; Conservative 7; Mismatches 23; Indels 0; Gaps 0;

QY 1 MLLKVAFAFAIVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQGSANNAALQ 60
 DB 1 MLLKVAFAFAIVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQGSANNAALQ 60
 QY 61 SDARKSETTITQSGYGADVGAGADNSTIELTONGFRNNATIDQWNAKNSDITVGYGG 120
 DB 61 SDARKSETTITQSGYGADVGAGADNSTIELTONGFRNNATIDQWNAKNSDITVGYGG 120
 QY 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151
 DB 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151

RESULT 14
 AAR62761
 ID AAR62761 standard; protein; 120 AA.
 XX
 AC AAR62761;
 XX
 DT 27-AUG-2003 (revised)
 DT 25-MAR-2003 (revised)
 DT 26-JUN-1995 (first entry)
 XX
 DE Agfa sequence.
 XX
 KW *Salmonella*, Agfa; vaccine.
 XX
 OS *Salmonella enteritidis*.
 OS
 PN WO9425598-A2.
 XX
 PD 10-NOV-1994.
 DR

XX 26-APR-1994; 94MO-IB000207.
 XX 26-APR-1993; 93US-00054452.
 XX (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
 PA (KING/) KING J.
 PI Kay WW, Collinson SK, Clouthier SC, Doran JL;
 XX WPI; 1994-358275/44.
 DR N-PSDB; AA073066.
 XX Eliciting an immune response to *Salmonella* - using attenuated *Salmonella*
 PT strains, vector constructs, or compens. contg. fimbrial type proteins.
 XX
 PS Disclosure; Fig 7A; 95pp; English.
 CC The sequence represents the *Salmonella enteritidis* 27655-3b *TnpH*Δ mutant
 CC strain Agfa protein. The encoding DNA and isolated Agfa protein are used
 CC in genetic immunization and vaccine compositions, respectively, to elicit
 CC an immune response to *Salmonella* in animals (e.g. food producing animals)
 CC and humans. (Updated on 25-MAR-2003 to correct PN field.) (Updated on 27-
 CC AUG-2003 to correct OS field.)
 SQ Sequence 120 AA;

Query Match 71.9%; Score 560; DB 2; Length 120;
 Best Local Similarity 100.0%; Pred. No. 1.2e-45;
 Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 VVPQWGGGNNHNGSGSPDSTLSIYOGSANAALALQSDARKSETTTTQSGYNGADV 81
 DB 1 VVPQWGGGNNHNGSGSPDSTLSIYOGSANAALALQSDARKSETTTTQSGYNGADV 60
 QY 82 GQGADNSTIELTONGFRNNATTIDQWNAKNSDITVGYGNNALVN 127
 DB 61 GQGADNSTIELTONGFRNNATTIDQWNAKNSDITVGYGNNALVN 106

RESULT 15

AAW23569
 ID AAW23569 standard; protein; 120 AA.

XX AAW23569;
 AC 25-MAR-2003 (revised)
 DT 29-SEP-1997 (first entry)
 XX
 DE *Salmonella enteritidis* 27655-3b *TnpH*Δ mutant agfa fragment.
 XX
 KM Enteropathogenic bacteria; enterobacteria; *S. enteritidis*; antibody.
 XX
 OS *Salmonella enteritidis*.
 XX
 PN US5635617-A.
 PD 03-JUN-1997.
 XX
 PF 26-APR-1994; 94US-00233788.
 XX
 PR 26-APR-1993; 93US-00054452.
 XX
 PA (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
 XX
 PI Collinson SK, Kay WW, Doran JL;
 XX
 DR WPI; 1997-309886/28.
 DR N-PSDB; AAT74141.
 XX
 PT Isolated *Salmonella* gene agfa - used for diagnosis of *Salmonella* or
 PT enteropathogenic bacteria of the Enterobacteria family.
 XX

PS Example 2; Fig 7; 85pp; English.

XX The present sequence represents an agfa fragment encoded by an agfa gene
 CC fragment derived from *Salmonella enteritidis* 27655-3b *TnpH*Δ mutant
 CC strain. The nucleic acid can be used to provide diagnostic assays for
 CC *Salmonella* and/or enteropathogenic bacteria of the family Enterobacteria.
 CC It can also be used to provide proteins and antibodies which can be used
 CC for assays. The nucleic acid sequence can be used to provide probes or
 CC primers which can specifically hybridise to nucleic acid molecules from
 CC greater than 99% of *Salmonella* strains that are pathogenic to warm-
 CC blooded animals relative to nucleic acid molecules from virtually all
 CC other microbial organisms. (Updated on 25-MAR-2003 to correct PF field.)
 XX

SQ Sequence 120 AA;

Query Match 71.9%; Score 560; DB 2; Length 120;
 Best Local Similarity 100.0%; Pred. No. 1.2e-45;
 Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 VVPQWGGGNNHNGSGSPDSTLSIYOGSANAALALQSDARKSETTTTQSGYNGADV 81
 DB 1 VVPQWGGGNNHNGSGSPDSTLSIYOGSANAALALQSDARKSETTTTQSGYNGADV 60
 QY 82 GQGADNSTIELTONGFRNNATTIDQWNAKNSDITVGYGNNALVN 127
 DB 61 GQGADNSTIELTONGFRNNATTIDQWNAKNSDITVGYGNNALVN 106

Search completed: March 11, 2004, 18:33:36
 Job time : 46.9 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:25:54 ; Search time 12.4 Seconds

(without alignments)
628,671 Million cell updates/sec

Title: US-09-543-407-12

Perfect score: 779

Sequence: 1 MLLKVAAPAAIVSGSALA.....VTRVTHMAHANNATANY 151

Scoring table: BLOSUM62

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgm2_6/prodata/2/iaa/5A.COMB.pep.*
2: /cgm2_6/prodata/2/iaa/5B.COMB.pep.*
3: /cgm2_6/prodata/2/iaa/6A.COMB.pep.*
4: /cgm2_6/prodata/2/iaa/6B.COMB.pep.*
5: /cgm2_6/prodata/2/iaa/PCTUS.COMB.pep.*
6: /cgm2_6/prodata/2/iaa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	630	88.6	151	US-08-233-788A-59	Sequence 59, App1
2	560	71.9	120	US-08-233-788A-57	Sequence 57, App1
3	560	71.9	120	US-09-196-387-8	Sequence 8, App1
4	92	11.8	673	US-09-841-835-8	Sequence 8, App1
5	92	11.8	949	US-09-196-387-10	Sequence 10, App1
6	92	11.8	949	US-09-841-835-10	Sequence 10, App1
7	92	11.8	1327	US-09-196-387-2	Sequence 2, App1
8	92	11.8	1327	US-09-841-835-2	Sequence 2, App1
9	92	11.8	1327	US-09-972-115A-8	Sequence 8, App1
10	90.5	11.6	745	US-09-336-115C-6	Sequence 6, App1
11	89.5	11.5	738	US-08-864-038A-3	Sequence 3, App1
12	89	11.4	943	US-09-056-556-204	Sequence 204, App
13	89	11.4	943	US-09-072-596-199	Sequence 199, App
14	89	11.4	943	US-09-477-135A-111	Sequence 131, App
15	89	11.4	943	US-09-072-967-204	Sequence 104, App
16	87.5	11.2	892	US-09-336-447A-5	Sequence 63, App1
17	85	10.9	956	US-09-134-078-63	Sequence 63, App1
18	82.5	10.6	186	US-09-382-276-3	Sequence 3, App1
19	82.5	10.6	873	US-09-336-447A-13	Sequence 13, App1
20	81.5	10.5	329	US-09-252-991A-32096	Sequence 32096, A
21	81	10.4	568	US-09-543-681A-6566	Sequence 6966, Ap
22	81	10.4	1739	US-09-540-236-3739	Sequence 3739, Ap
23	81	10.4	1864	US-08-804-227C-3	Sequence 3, App1
24	80	10.3	232	US-09-252-991A-30263	Sequence 30263, A
25	80	10.3	975	US-09-328-352-4764	Sequence 4764, Ap
26	80	10.3	3241	US-09-841-786-1	Sequence 1, App1
27	79.5	10.2	361	US-09-540-236-2164	Sequence 2164, Ap

28	79	10.1	941	US-09-336-447A-9	Sequence 9, App1
29	78.5	10.1	2123	US-08-968-685A-10	Sequence 10, App1
30	78	10.0	906	US-08-254-573-2	Sequence 2, App1
31	78	10.0	906	US-08-687-379-2	Sequence 2, App1
32	78	10.0	906	US-08-687-379-4	Sequence 4, App1
33	78	10.0	906	US-08-172-332-1	Sequence 1, App1
34	78	10.0	906	US-08-216-326-2	Sequence 2, App1
35	77.5	9.9	702	US-09-252-991A-22119	Sequence 22119, A
36	77.5	9.9	714	US-09-841-786-4	Sequence 4, App1
37	77.5	9.9	878	US-09-540-236-3401	Sequence 3401, Ap
38	77	9.9	415	US-09-025-769B-280	Sequence 280, App
39	77	9.9	528	US-09-490-291-8	Sequence 8, App1
40	77	9.9	1690	US-09-595-684B-39	Sequence 39, App1
41	76.5	9.8	2315	US-09-543-681A-5434	Sequence 5434, Ap
42	76.5	9.8	159	US-08-856-253-2	Sequence 2, App1
43	76.5	9.8	186	US-09-382-276-2	Sequence 2, App1
44	76.5	9.8	605	US-09-489-039A-13002	Sequence 13002, A
45	75.5	9.7	624	US-09-336-447A-7	Sequence 7, App1

ALIGNMENTS

RESULT 1
US-08-233-788A-59
Sequence 59, Application US/08233788A
Patent No. 5635617
GENERAL INFORMATION:
APPLICANT: Doran, James L.
APPLICANT: Kay, William W.
APPLICANT: Collinson, Karen S.
APPLICANT: Clouthier, Sharon C.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
NUMBER OF INVENTION: OF SALMONELLA
CORRESPONDENCE ADDRESS: 61
ADDRESSER: Seed and Berry
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: U.S.A.
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233,788A
FILING DATE: 26-APR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: King, Joshua
REGISTRATION NUMBER: 35,570
REFERENCE/DOCKET NUMBER: 920043.403C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEDANDBERY
INFORMATION FOR SEQ ID NO: 59:
SEQUENCE CHARACTERISTICS:
LENGTH: 151 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-233-788A-59

Query Match 88.6%; Score 690; DB 1; Length 151;
Best Local Similarity 89.4%; Pred. No. 7.1e-62;
Matches 135; Conservative 4; Mismatches 12; Indels 0; Gaps 0;
1 MLLKVAAPAAIVSGSALAGVPPWGGGNGGNGSSGPGSTISITYGSNAAALAQ 60
|||||

Db 1 MLLKVAFAAIVVSSGALAGVVPQWGGGNNNGGNSGPDSTLSIYOGSANAALALQ 60
 QY 61 SDARKSETTTTQSGYNGADVGQADNSTIELTONGFRNNATIDQNNAKNSDIYGOYGG 120
 Db 61 SDARKSETTTTQSGYNGADVGQADNSTIELTONGFRNNATIDQNNAKNSDIYGOYGG 120
 QY 121 NNAALVYDQLVRYVTHMAHANNATANOY 151
 Db 121 NNPALVYQNTASDSSVWQVGFQNNATANOY 151

RESULT 2

US-08-233-788A-57
 ; Sequence 57, Application US/08233788A
 ; Patent No. 5635617

GENERAL INFORMATION:

APPLICANT: Doran, James L.
 APPLICANT: Kay, William W.
 APPLICANT: Collinson, Karen S.
 APPLICANT: Clouthier, Sharon C.
 TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
 NUMBER OF SEQUENCES: 61
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Seed and Berry
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington
 COUNTRY: U.S.A.
 ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/233,788A
 FILING DATE: 26-APR-1994
 CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: King, Joshua
 REGISTRATION NUMBER: 35,570
 REFERENCE/DOCKET NUMBER: 920043.403C2
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 622-4900
 TELEFAX: (206) 682-6031
 TELEX: 3723836 SEDANBERRY

INFORMATION FOR SEQ ID NO: 57:

SEQUENCE CHARACTERISTICS:
 LENGTH: 120 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-233-788A-57

Query Match 71.9%; Score 560; DB 1; Length 120;
 Best Local Similarity 100.0%; Pred. No. 66-49;
 Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 VVPQWGGGNNNGGNSGPDSTLSIYOGSANAALALQSDARKSETTTTQSGYNGADV 81
 Db 1 VVPQWGGGNNNGGNSGPDSTLSIYOGSANAALALQSDARKSETTTTQSGYNGADV 60
 QY 82 GQAGDNSTIELTONGFRNNATIDQNNAKNSDIYGOYGGNNAALVN 127
 Db 61 GQAGDNSTIELTONGFRNNATIDQNNAKNSDIYGOYGGNNAALVN 106

RESULT 3

US-09-196-387-8
 ; Sequence 8, Application US/09196387
 ; Patent No. 6277613
 ; GENERAL INFORMATION:

APPLICANT: de Lange, Titia
 APPLICANT: Smith, Susan
 TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
 TITLE OF INVENTION: OF USE THEREOF
 NUMBER OF SEQUENCES: 12
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Klauber & Jackson
 STREET: 411 Hackensack Avenue, 4th Floor
 CITY: Hackensack
 STATE: New Jersey
 COUNTRY: USA
 ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/196,387
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 09/095,225
 FILING DATE: June 10, 1998
 ATTORNEY/AGENT INFORMATION:
 NAME: Jackson Esq., David A.
 REGISTRATION NUMBER: 26,742
 REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 201-487-5800
 TELEFAX: 201-343-1684
 TELEX: 133521

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:
 LENGTH: 673 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-09-196-387-8

Query Match 11.8%; Score 92; DB 3; Length 673;
 Best Local Similarity 30.4%; Pred. No. 0.6;
 Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

QY 6 VAAFAAI-VSSGALAGVVPQWGGGNNNGGNSGPDSTLSIYOGSANAALALQSDAR 64
 Db 99 VAAFPVPAVSTSSAAGVAPNPAGSGNSPSSSSPTSS-SSSSPSSPGSSIAESPEAA 157
 QY 65 KSETTIT---QSGYNGADVGQADNSTIELTONG--FRNNATIDQNNAKNSDI 113
 Db 158 GVSSTAPLPGPMAAGPTGVPAVSGALRELLACRNGDVSRVKRLVDAAVNAKDM 212

RESULT 4

US-09-841-835-8
 ; Sequence 8, Application US/09841835
 ; Patent No. 6506587
 ; GENERAL INFORMATION:

APPLICANT: de Lange, Titia
 APPLICANT: Smith, Susan
 TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
 TITLE OF INVENTION: OF USE THEREOF
 NUMBER OF SEQUENCES: 12
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Klauber & Jackson
 STREET: 411 Hackensack Avenue, 4th Floor
 CITY: Hackensack
 STATE: New Jersey
 COUNTRY: USA
 ZIP: 07601

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible

Db 99 VAAAPVPAVSTSSAAGVAPNPAGSGSNNSPSSSSPTSS-SSSSPSSPGSSLAESPFA 157
 QY 65 KSETTTT---OSGYNGADVGGAGNNTIELTONG--FRNNATTIDOWNAKNSDI 113
 Db 158 GVSSTAPLPGAGAGGTGTGVPVAVSGALRELLACRNGDVSRVRLVDAAVNAKDM 212

RESULT 7

US-09-196-387-2
 ; Sequence 2, Application US/09196387
 ; Patent No. 6277613

GENERAL INFORMATION:
 APPLICANT: de Lange, Titia
 APPLICANT: Smith, Susan
 TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRP1 AND METHODS
 TITLE OF INVENTION: OF USE THEREOF
 NUMBER OF SEQUENCES: 12
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Klauber & Jackson
 STREET: 411 Hackensack Avenue, 4th Floor
 CITY: Hackensack
 STATE: New Jersey
 COUNTRY: USA
 ZIP: 07601

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/196,387
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 09/095,225
 FILING DATE: June 10, 1998

ATTORNEY/AGENT INFORMATION:
 NAME: Jackson Esq., David A.
 REGISTRATION NUMBER: 26,742
 REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 201-487-5800
 TELEFAX: 201-343-1684
 TELEX: 133521

INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1327 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHETICAL: NO
 US-09-196-387-2

Query Match 11.8%; Score 92; DB 3; Length 1327;
 Best Local Similarity 30.4%; Pred. No. 1.5;
 Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

QY 6 VAAFAAI-VVSGSALAGVVPQWGGGNNNGGNSGPDSTLSIYQGSANAAALAQSDAR 64
 Db 99 VAAAPVPAVSTSSAAGVAPNPAGSGSNNSPSSSSPTSS-SSSSPSSPGSSLAESPFA 157
 QY 65 KSETTTT---OSGYNGADVGGAGNNTIELTONG--FRNNATTIDOWNAKNSDI 113
 Db 158 GVSSTAPLPGAGAGGTGTGVPVAVSGALRELLACRNGDVSRVRLVDAAVNAKDM 212

RESULT 8
 US-09-841-835-2
 ; Sequence 2, Application US/09841835
 ; Patent No. 6506587
 ; GENERAL INFORMATION:

APPLICANT: de Lange, Titia
 APPLICANT: Smith, Susan
 TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRP1 AND METHODS
 TITLE OF INVENTION: OF USE THEREOF
 NUMBER OF SEQUENCES: 12
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Klauber & Jackson
 STREET: 411 Hackensack Avenue, 4th Floor
 CITY: Hackensack
 STATE: New Jersey
 COUNTRY: USA
 ZIP: 07601

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/841,835
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 09/196,387
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Jackson Esq., David A.
 REGISTRATION NUMBER: 26,742
 REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 201-487-5800
 TELEFAX: 201-343-1684
 TELEX: 133521

INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1327 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHETICAL: NO
 US-09-841-835-2

Query Match 11.8%; Score 92; DB 4; Length 1327;
 Best Local Similarity 30.4%; Pred. No. 1.5;
 Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

QY 6 VAAFAAI-VVSGSALAGVVPQWGGGNNNGGNSGPDSTLSIYQGSANAAALAQSDAR 64
 Db 99 VAAAPVPAVSTSSAAGVAPNPAGSGSNNSPSSSSPTSS-SSSSPSSPGSSLAESPFA 157
 QY 65 KSETTTT---OSGYNGADVGGAGNNTIELTONG--FRNNATTIDOWNAKNSDI 113
 Db 158 GVSSTAPLPGAGAGGTGTGVPVAVSGALRELLACRNGDVSRVRLVDAAVNAKDM 212

RESULT 9
 US-09-972-115A-8
 ; Sequence 8, Application US/09972115A
 ; Patent No. 6599728
 ; GENERAL INFORMATION:
 APPLICANT: Geron Corporation
 APPLICANT: Gregg, Morin B.
 APPLICANT: Walter, Funk D.
 APPLICANT: Mieczyslaw, Piatyszek A.
 TITLE OF INVENTION: A Second Mammalian Telomerase
 FILE REFERENCE: 080/003C
 CURRENT APPLICATION NUMBER: US/09/972,115A
 PRIOR FILING DATE: 2001-10-05
 PRIOR APPLICATION NUMBER: US 60/128,577
 PRIOR FILING DATE: 2000-04-10
 PRIOR APPLICATION NUMBER: US 60/129,123
 PRIOR FILING DATE: 1999-04-13
 NUMBER OF SEQ ID NOS: 64


```

ADDRESSER: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/056,556
FILING DATE: 07-APR-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: MAKI, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.457
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 204:
SEQUENCE CHARACTERISTICS:
LENGTH: 943 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-09-056-556-204

Query Match
Best Local Similarity 11.4%; Score 89; DB 4; Length 943;
Matches 33; Conservative 15; Mismatches 55; Indels 26; Gaps 5;

QY 12 IVVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQSDAR---KSET 68
DB 159 IGLTSGGLGF-----GGLNSGTGN-----IGLFNSGTGNVIGNSGTGNMGIGNSG 205
QY 69 TTOSGYNGADVQGGADNSTIELTONGFRNNATTIDQWNAKNSD---ITVGOY----- 118
DB 206 NSYNTGFGNSGDANTGFRFNSGIANTGVNAGNYNTGNSYNTGNSGTGFMGQYNTGYLNS 265
QY 119 GGNNAALVN 127
DB 266 GNYNTGLAN 274

RESULT 13
US-09-072-596-199
Sequence 199, Application US/09072596
Patent No. 6458366
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonia
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel J.
APPLICANT: Lodes, Michael J.
APPLICANT: Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 350
CORRESPONDENCE ADDRESS:
ADDRESSER: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

```

```

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/072,596
FILING DATE: 05-MAY-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: MAKI, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 199:
SEQUENCE CHARACTERISTICS:
LENGTH: 943 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-09-072-596-199

Query Match
Best Local Similarity 11.4%; Score 89; DB 4; Length 943;
Matches 33; Conservative 15; Mismatches 55; Indels 26; Gaps 5;

QY 12 IVVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQSDAR---KSET 68
DB 159 IGLTSGGLGF-----GGLNSGTGN-----IGLFNSGTGNVIGNSGTGNMGIGNSG 205
QY 69 TTOSGYNGADVQGGADNSTIELTONGFRNNATTIDQWNAKNSD---ITVGOY----- 118
DB 206 NSYNTGFGNSGDANTGFRFNSGIANTGVNAGNYNTGNSYNTGNSGTGFMGQYNTGYLNS 265
QY 119 GGNNAALVN 127
DB 266 GNYNTGLAN 274

RESULT 14
US-09-477-135A-131
Sequence 131, Application US/09477135A
Patent No. 6572865
GENERAL INFORMATION:
APPLICANT: Nano, Francis
TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
FILE REFERENCE: 52888
CURRENT APPLICATION NUMBER: US/09/477,135A
CURRENT FILING DATE: 2000-01-03
PRIOR APPLICATION NUMBER: 08990823
PRIOR FILING DATE: 1997-12-15
PRIOR APPLICATION NUMBER: US 96/10375
PRIOR FILING DATE: 1996-06-14
PRIOR APPLICATION NUMBER: 60/000,254
PRIOR FILING DATE: 1995-06-15
NUMBER OF SEQ ID NOS: 169
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 131
LENGTH: 943
TYPE: PRT
ORGANISM: Mycobacterium tuberculosis
US-09-477-135A-131

Query Match
Best Local Similarity 11.4%; Score 89; DB 4; Length 943;
Matches 33; Conservative 15; Mismatches 55; Indels 26; Gaps 5;

QY 12 IVVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQSDAR---KSET 68
DB 266 IGLTSGGLGF-----GGLNSGTGN-----IGLFNSGTGNVIGNSGTGNMGIGNSG 312
QY 69 TTOSGYNGADVQGGADNSTIELTONGFRNNATTIDQWNAKNSD---ITVGOY----- 118

```


THIS PAGE BLANK (USPTO)

NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742

Query Match	11.8%;	Score 92;	DB 9;	Length 949;
Best Local Similarity	30.4%;	Pred. No. 1.8;		

Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

QY 6 VAAFAAI-VVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSTIYQGSANAALALQSDAR 64

DB 99 VAAAPVPAVSTSSAAGVAPNPAGSGSNNSSPSSSPTSS-SSSSPSSPSSSLAESPEAA 157

QY 65 KSETTIT---OSGYNGADVGGADNSTIELTONG--FRNNATTIDQWAKNSDI 113

DB 158 GVSSTAPLPGGAAGPCTGVPAVSGALRELLACRNGDVSRRVRLVDAANVNAKDM 212

RESULT 5

US-09-841-835-2

Sequence 2, Application US/09841835

Patent No. US20020076795A1

GENERAL INFORMATION:

APPLICANT: de Lange, Titia

APPLICANT: Smith, Susan

TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS

TITLE OF INVENTION: OF USE THEREOF

CORRESPONDENCE ADDRESS:

ADDRESSEE: Klauber & Jackson

STREET: 411 Hackensack Avenue, 4th Floor

CITY: Hackensack

STATE: New Jersey

COUNTRY: USA

ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/841,835

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/196,387

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Jackson Esq., David A.

REGISTRATION NUMBER: 26,742

REFERENCE/DOCKET NUMBER: 600-1-230 CIP1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201-487-5800

TELEFAX: 201-343-1684

TELEX: 133521

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1327 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULAR TYPE: protein

HYPOTHETICAL: NO

US-09-841-835-2

Query Match

Best Local Similarity 11.8%; Score 92; DB 9; Length 1327;

Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

QY 6 VAAFAAI-VVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSTIYQGSANAALALQSDAR 64

DB 99 VAAAPVPAVSTSSAAGVAPNPAGSGSNNSSPSSSPTSS-SSSSPSSPSSSLAESPEAA 157

QY 65 KSETTIT---OSGYNGADVGGADNSTIELTONG--FRNNATTIDQWAKNSDI 113

DB 158 GVSSTAPLPGGAAGPCTGVPAVSGALRELLACRNGDVSRRVRLVDAANVNAKDM 212

RESULT 6

US-09-972-115A-8

Sequence 8, Application US/09972115A

Publication No. US20030032769A1

GENERAL INFORMATION:

APPLICANT: Geron Corporation

APPLICANT: Gregg, Morin B.

APPLICANT: Walter, Funk D.

APPLICANT: Mieczyslaw, Piatyszek A.

TITLE OF INVENTION: A Second Mammalian Telomerase

FILE REFERENCE: 080/003C

CURRENT APPLICATION NUMBER: US/09/972,115A

CURRENT FILING DATE: 2001-10-05

PRIOR APPLICATION NUMBER: US 60/128,577

PRIOR FILING DATE: 2000-04-10

PRIOR APPLICATION NUMBER: US 60/129,123

PRIOR FILING DATE: 1999-04-13

NUMBER OF SEQ ID NOS: 64

SOFTWARE: Patent in version 3.1

SEQ ID NO 8

LENGTH: 1327

TYPE: PRT

ORGANISM: Homo sapiens

US-09-972-115A-8

Query Match

Best Local Similarity 11.8%; Score 92; DB 10; Length 1327;

Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

QY 6 VAAFAAI-VVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSTIYQGSANAALALQSDAR 64

DB 99 VAAAPVPAVSTSSAAGVAPNPAGSGSNNSSPSSSPTSS-SSSSPSSPSSSLAESPEAA 157

QY 65 KSETTIT---OSGYNGADVGGADNSTIELTONG--FRNNATTIDQWAKNSDI 113

DB 158 GVSSTAPLPGGAAGPCTGVPAVSGALRELLACRNGDVSRRVRLVDAANVNAKDM 212

RESULT 7

US-10-199-937-4

Sequence 4, Application US/10199937

Publication No. US20030190739A1

GENERAL INFORMATION:

APPLICANT: Christenson, Erik

APPLICANT: Demaggio, Anthony J.

APPLICANT: Goldman, Phyllis S.

APPLICANT: McElligott, David L.

TITLE OF INVENTION: TANKYRASE2 MATERIALS AND METHODS

FILE REFERENCE: 27866/36559

CURRENT APPLICATION NUMBER: US/10/199,937

CURRENT FILING DATE: 2002-07-22

PRIOR APPLICATION NUMBER: US/09/606,035

PRIOR FILING DATE: 2000-06-28

PRIOR APPLICATION NUMBER: 60/141,582

PRIOR FILING DATE: 1999-06-29

NUMBER OF SEQ ID NOS: 178

SOFTWARE: Patent in Ver. 2.0

SEQ ID NO 4

LENGTH: 1327

TYPE: PRT

ORGANISM: Homo sapiens

US-10-199-937-4

Query Match

Best Local Similarity 11.8%; Score 92; DB 14; Length 1327;

Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

QY 6 VAAFAAI-VVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSTIYQGSANAALALQSDAR 64

DB 99 VAAAPVPAVSTSSAAGVAPNPAGSGSNNSSPSSSPTSS-SSSSPSSPSSSLAESPEAA 157

QY 65 KSETTIT---OSGYNGADVGGADNSTIELTONG--FRNNATTIDQWAKNSDI 113

DB 158 GVSSTAPLPGGAAGPCTGVPAVSGALRELLACRNGDVSRRVRLVDAANVNAKDM 212

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:24:14; Search time 10.3 Seconds
(without alignment)
1410.186 Million cell updates/sec

Title: US-09-543-407-12

Perfect score: 779
Sequence: 1 MKLLKVAFAFAIVSGSALA.....VTRVVTHEMAHANNATANOY 151

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: PIR 78:*

1: PIR1: *
2: PIR2: *
3: PIR3: *
4: PIR4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	695	89.2	151	2	JC6039
2	695	89.2	151	2	AT0635
3	520	66.8	151	2	S70788
4	501.5	64.4	152	2	D90806
5	501.5	64.4	152	2	H85665
6	106	13.6	1748	2	S42136
7	104	13.4	2174	2	E95965
8	101.5	13.0	151	2	AH0635
9	100.5	12.9	151	2	UC6040
10	98.5	12.6	1028	2	A56038
11	98.5	12.6	1213	2	S16356
12	96	12.3	145	2	AD3143
13	96	12.3	145	2	H98144
14	94.5	12.1	678	2	A70762
15	93.5	12.0	1238	2	AH0038
16	93	11.9	573	2	C86366
17	91.5	11.7	151	2	S70787
18	91.5	11.7	151	2	C90806
19	91.5	11.7	151	2	C85665
20	91.5	11.7	256	2	T03371
21	91.5	11.7	770	2	T51024
22	91	11.7	1250	2	D91018
23	91	11.7	1250	2	P85862
24	90.5	11.6	745	2	E64559
25	90	11.6	409	2	T20847
26	90	11.6	575	2	S35327
27	89.5	11.5	582	2	F70675
28	89.5	11.5	262	2	S00275
29	89.5	11.5	343	2	T05221

30	89	11.4	347	2	B39112	merozoite 45K surf
31	89	11.4	3300	2	D70575	probable PPE prote
32	88.5	11.4	407	2	T21956	hypothetical prote
33	88.5	11.4	447	2	G84687	probable disease r
34	88.5	11.4	552	2	D70604	probable PPE prote
35	88	11.3	301	2	B84533	hypothetical prote
36	88	11.3	354	2	B70663	probable PPE prote
37	88	11.3	434	2	E70768	hypothetical glyci
38	87.5	11.2	615	2	E70663	probable PPE prote
39	87.5	11.2	940	2	D89723	protein F39D8.1b l
40	87.5	11.2	945	2	T21998	hypothetical prote
41	87	11.2	331	2	S21409	class 3 outer memb
42	87	11.2	440	2	AD1539	probable sugar ABC
43	87	11.2	645	2	F70825	probable PPE prote
44	86.5	11.1	1635	2	A10452	hemolysin [impor
45	86.5	11.1	3190	2	T13828	CREB-binding prote

ALIGNMENTS

RESULT 1

JC6039 fimbriin protein agfa precursor - Salmonella enteritidis

C/Species: Salmonella enteritidis

C/Date: 31-Dec-1996 #sequence_revision 31-Dec-1996 #text_change 08-Oct-1999

C/Accession: JC6039; PC6015; A44898

R/Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Baner, P.A.; Kay, W.W.

J. Bacteriol. 178, 662-667, 1996

A/Title: Salmonella enteritidis agfA operon encoding thin, aggregative fimbriae.

A/Reference number: JC6039; MUID:96146512; PMID:8550497

A/Accession: JC6039

A/Molecule type: DNA

A/Residues: 1-151 <COL>

A/Cross-references: GB:U43280; NID:91184712; PIDN:ACC43599.1; PID:91184714

A/Accession: PC6015

A/Molecule type: protein

A/Residues: 21-52 <CO2>

A/Experimental source: strain 27655-3b

A/Note: the authors translated the codon ACG for residue 44 as Ile

R/Collinson, S.K.; Emody, L.; Muller, K.H.; Trust, T.J.; Kay, W.W.

J. Bacteriol. 173, 4773-4781, 1991

A/Title: Purification and characterization of thin, aggregative fimbriae from Salmonell.

A/Reference number: A44898; MUID:91310586; PMID:1677357

A/Accession: 27655

A/Contents: A44898

A/Status: preliminary

A/Molecule type: protein

A/Residues: 21-33 <CO3>

A/Note: sequence extracted from NCBI backbone (NCBIP:45936)

C/GeneticB

A/Genes: agfA

C/Function:

A/Description: major component of thin aggregative fimbriae

A/Note: fimbriae bind to fibronectin, plasminogen, tissue plasminogen activator

F.1-30/Domain: signal sequence #status predicted <SIG>

F.1-151/Product: fimbriin protein agfA #status experimental <MAT>

Query Match

Best Local Similarity 90.1%; Pred. No. 1.8e-50;

Matches 136; Conservative 4; Mismatches 11; Indels 0; Gaps 0;

Query

1 MKLLKVAFAFAIVSGSALAGVVPQWGGGNNHGGSSGPDSTLSTIYGSAANAALQ 60

1 MKLLKVAFAFAIVSGSALAGVVPQWGGGNNHGGSSGPDSTLSTIYGSAANAALQ 60

61 SDARKSETTTTQSGYGADVGADNSTLELTONGFRNNAITDQNAKSDITVGGYGG 120

61 SDARKSETTTTQSGYGADVGADNSTLELTONGFRNNAITDQNAKSDITVGGYGG 120

121 NNAAALVNYDQVTRVVTHEMAHANNATANOY 151

121 NNAAALVNYDQVTRVVTHEMAHANNATANOY 151

Db 121 NNAALVNOTASDSSVWVQVGFNNATANY 151

RESULT 2

A:0635

major curlin chain precursor [imported] - *Salmonella enterica* subsp. *enterica* serovar Typhimurium
C/Species: *Salmonella enterica* subsp. *enterica* serovar Typhimurium
C/Note: this species has also been called *Salmonella typhimurium*
C/Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C/Accession: A10635
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Conerton, F.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Garra, P.
A:Authors: Parky, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhimurium
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: A10635
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-151 <PAR>
C/Cross-references: GB:AL513382; PIDN:CAM08268.1; PID:g16502315; GSPDB:GN00176
C/Genetics:
A:Gene: STM1181

Query Match 89.2%; Score 695; DB 2; Length 151;
Best Local Similarity 90.1%; Pred. No. 1.8e-50;
Matches 136; Conservative 4; Mismatches 11; Indels 0; Gaps 0;

QY 1 MKLLKVAAPALIVSSGALAGVPPQWGGGNNNGSGSPDSTLSIYOGSANAALALQ 60
Db 1 MKLLKVAAPALIVSSGALAGVPPQWGGGNNNGSGSPDSTLSIYOGSANAALALQ 60
QY 61 SPARKSETTITGSGYNGADVGAGADNSTITETLTONGFNNATITDQMNKNSDITVGOY 120
Db 61 SPARKSETTITGSGYNGADVGAGADNSTITETLTONGFNNATITDQMNKNSDITVGOY 120
QY 121 NNAALVNOTASDSSVWVQVGFNNATANY 151
Db 121 NNAALVNOTASDSSVWVQVGFNNATANY 151

RESULT 3

S70788

curlin protein csgA precursor - *Escherichia coli* (strain K-12)
N/Alternate names: csgA protein; major curlin protein
C/Species: *Escherichia coli*
C/Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 01-Mar-2002
C/Accession: S70788; G64846; S31202; S34560; S34559
R:Hammar, M.; Argyris, A.; Bian, Z.; Olsen, A.; Normark, S.
Mol. Microbiol. 18, 661-670, 1995
A:Title: Expression of two csg operons is required for production of fibronectin- and Cc
A:Reference number: S70783; MUID:96414468; PMID:8817489
A:Accession: S70788
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-151 <HAM>

A:Cross-references: EMBL:X90754; NID:g1147558; PIDN:CAA62282.1; PID:g1147564
A:Experimental source: strain K12, substrain W3110
A:Note: The nucleotide sequence was submitted to the EMBL Data Library, August 1995
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
A:Authors: D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of *Escherichia coli* K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: G64846
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-151 <BIAT>

A:Cross-references: GB:AE000205; GB:U00096; NID:g1787265; PIDN:AACT4126.1; PID:g1787279;
R:Olsen, A.; Argyris, A.; Hammar, M.; Sukupoly, S.; Normark, S.
Mol. Microbiol. 7, 523-536, 1993

A:Title: The Rpos sigma factor relieves H-NS-mediated transcriptional repression of csgA
A:Reference number: S31202; MUID:93211294; PMID:8459772
A:Accession: S31202

A:Molecule type: DNA
A:Residues: 1-6, 'v', 8-151 <OL51>
A:Cross-references: EMBL:U04979
A:Accession: S34560
A:Molecule type: protein
A:Residues: 21-42; 44-50 <OL52>
R:Olsen, A.N.; Argyris, A.M.
submitted to the EMBL Data Library, October 1992
A:Reference number: S34559
A:Accession: S34559

A:Molecule type: DNA
A:Residues: 1-133, 'RQRDSGWLW' <OL53>
A:Cross-references: EMBL:U04979; NID:g290424; PIDN:AAA23616.1; PID:g290425
A:Experimental source: strain K-12, substrain W3110
C/Genetics:
A:Gene: csgA
A:Map position: 23.15
A:Function:
A:Description: major component of wild-type curli; interaction between CsgA and CsgB to
A:Note: curli are thin, coiled fibers expressed on the surface of *Escherichia coli* that
F1-20/Domain: signal sequence #status predicted <Sig>
F121-151/Product: curlin #status experimental <Mat>

Query Match 66.8%; Score 520; DB 2; Length 151;
Best Local Similarity 66.9%; Pred. No. 4.7e-36;
Matches 101; Conservative 21; Mismatches 29; Indels 0; Gaps 0;

QY 1 MKLLKVAAPALIVSSGALAGVPPQWGGGNNNGSGSPDSTLSIYOGSANAALALQ 60
Db 1 MKLLKVAAPALIVSSGALAGVPPQWGGGNNNGSGSPDSTLSIYOGSANAALALQ 60
QY 61 SPARKSETTITGSGYNGADVGAGADNSTITETLTONGFNNATITDQMNKNSDITVGOY 120
Db 61 SPARKSETTITGSGYNGADVGAGADNSTITETLTONGFNNATITDQMNKNSDITVGOY 120
QY 121 NNAALVNOTASDSSVWVQVGFNNATANY 151
Db 121 NNAALVNOTASDSSVWVQVGFNNATANY 151

RESULT 4

D90806

curlin major subunit CsgA [imported] - *Escherichia coli* (strain O157:H7, substrain RIMD
D90806
C/Species: *Escherichia coli*
C/Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C/Accession: D90806
R:Hayashi, T.; Makino, K.; Onishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.C
gasaawa, N.; Yasunaga, T.; Kuwara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and gen
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: D90806
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-152 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA34843.1; PID:g13360880; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C/Genetics:
A:Gene: EcsA1420

Query Match 64.4%; Score 501.5; DB 2; Length 152;
Best Local Similarity 65.8%; Pred. No. 1.6e-34;
Matches 100; Conservative 21; Mismatches 30; Indels 1; Gaps 1;

QY 1 MKLLKVAAPALIVSSGALAGVPPQWGGGNNNGSGSPDSTLSIYOGSANAALALQ 59
Db 1 MKLLKVAAPALIVSSGALAGVPPQWGGGNNNGSGSPDSTLSIYOGSANAALALQ 60
QY 60 QSDARKSETTITGSGYNGADVGAGADNSTITETLTONGFNNATITDQMNKNSDITVGOY 119

```

Db      61 QADNRNSDLTTTGGGNGADVGGSGDDSSIDLTRQFGNSATLDDQNGKDSHTVTVQFG 120
QY      120 GNNALVNYDQLVTRVVTHEMAHANNATANQY 151
      121 GGNAAVDDQTAASNTVAVTVQVFGNNATANQY 152

RESULT 5
H85665
hypothetical protein csgA [imported] - Escherichia coli (strain O157:H7, substrain EDL93
C/Species: Escherichia coli
C/Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C/Accession: H85665
R/Perma, N.T.: Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A/Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A/Reference number: A85480; MUID:21074935; PMID:11206551
A/Accession: H85665
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-152 <STO>
A/Cross-references: GB:AE005174; NID:g12514574; PIDN:AA65788.1; GSPDB:GN00145; UMCP:216
A/Experimental source: strain O157:H7, substrain EDL933
A/Genes: csgA

Query Match      64.4%; Score 501.5; DB 2; Length 152;
Best Local Similarity 65.8%; Pred. No. 1.6e-34;
Matches 100; Conservative 21; Mismatches 30; Indels 1; Gaps 1;

QY      1 MLKLKVAFAAIIVSGSALAGVVPQW-GGGGNHNGGSSGPDSTLSIYOGSANAALAL 59
      1 MLKLKVAFAAIIVFSSGALAGVVPQYGGGNGGSGNNGPNSLNTIYOGGNSALAL 60
DB      1
QY      60 QSDARSETTTTSGGNGADVGGGADNSTIELTONGFRNNATIDQNAKNSDITVQYQ 119
      61 QADNRNSDLTTTGGGNGADVGGSGDDSSIDLTRQFGNSATLDDQNGKDSHTVTVQFG 120
DB      120 GNNALVNYDQLVTRVVTHEMAHANNATANQY 151
      121 GGNAAVDDQTAASNTVAVTVQVFGNNATANQY 152
DB      121

RESULT 6
S42136
cnbB protein - Tetrahymena thermophila
C/Species: Tetrahymena thermophila
C/Date: 19-Mar-1997 #sequence_revision 18-Jul-1997 #text_change 07-Dec-1999
C/Accession: S42136; S42135; S03650
R/Taylor, F.M.; Martindale, D.W.
submitted to the EMBL Data Library, October 1992
A/Reference number: S42136
A/Accession: S42136
A/Molecule type: DNA
A/Residues: 1-1748 <TAB>
A/Cross-references: EMBL:L03710; NID:g161751; PID:g161752
R/Taylor, F.M.; Martindale, D.W.
Nucleic Acids Res. 21, 4610-4614, 1993
A/Title: Retroviral-type zinc fingers and glycine-rich repeats in a protein encoded by c
A/Reference number: S42135; MUID:94051569; PMID:8233798
A/Accession: S42135
A/Molecule type: DNA
A/Residues: 1164-1174; 1179-1198; 1233-1252; 1285-1293; 1297-1309; 1316-1326; 1331-1341; 1343-1
A/Cross-references: EMBL:L03710
R/Martindale, D.W.; Taylor, F.M.
Nucleic Acids Res. 16, 2189-2201, 1988
A/Title: Multiple functions in a conjugation-specific gene from Tetrahymena thermophila.
A/Reference number: S03650; MUID:8819811; PMID:3357771
A/Accession: S03650
A/Molecule type: DNA
A/Residues: 236-250; T', 252-255; N', 257-773 <MAR>

```

```

A/Cross-references: EMBL:X06462
A/Genes: cnbB
A/Genetic code: SGCS
A/Intons: 85/3; 136/1; 157/3; 201/2; 290/2; 327/3; 499/1; 573/2; 607/3; 708/3; 777/3;
C/Keywords: zinc finger
F/1164-1450/Region: zinc finger
F/1451-1464/Region: zinc finger CCHC motif
F/1478-1491/Region: zinc finger CCHC motif
F/1501-1514/Region: zinc finger CCHC motif
F/1530-1543/Region: zinc finger CCHC motif
F/1555-1568/Region: zinc finger CCHC motif
F/1579-1592/Region: zinc finger CCHC motif
F/1602-1615/Region: zinc finger CCHC motif
F/1626-1748/Region: glycine-rich

Query Match      13.6%; Score 106; DB 2; Length 1748;
Best Local Similarity 32.1%; Pred. No. 1;
Matches 35; Conservative 13; Mismatches 33; Indels 28; Gaps 5;

QY      25 QMGGGNHNGG---GNSSGPDSTLSIYOGSANAALALQSDARKSETTIF---QSGYGN 77
      1640 QFGGGNSNGGSGWGTSSGSDNN-----CQSNVSESTTSSGGMGSSGSGN 1685
DB      1640
QY      78 GADVGGADNSTIELTONGFRNNATIDQNAKNSDITVQY-QYGNNA 124
      1686 QFGGGNSNDN-----QOQONENTGGGCGSSNSQNTNNSSGNNQA 1729
DB      1686

```

```

RESULT 7
E95965
hypothetical glycine-rich protein [imported] - Sinorhizobium meliloti (strain 1021) mag
C/Species: Sinorhizobium meliloti
C/Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C/Accession: E95965
R/Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Herna
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A/Title: The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing end
A/Reference number: A95842; MUID:21396508; PMID:11481431
A/Accession: E95965
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-2174 <KUR>
A/Cross-references: GB:AL51985; PIDN:CAC49389.1; PID:g15140875; GSPDB:GN00167
A/Experimental source: strain 1021, megaplasmid pSymB
R/Galbert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
Pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A/Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kise, E.; Komp, C.; Leleure
heault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K
A/Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A/Reference number: A96039; MUID:21368234; PMID:11474104
A/Contents: annotation
A/Genes: Smb21548
A/Genome: plasmid

Query Match      13.4%; Score 104; DB 2; Length 2174;
Best Local Similarity 27.1%; Pred. No. 1.9;
Matches 35; Conservative 16; Mismatches 44; Indels 34; Gaps 6;

QY      11 ATIVSGSALAGVVPQ-WGGGNGHNGGSSGPDSTLSIYOGSANAALALQSDARKSET 68
      693 AATATGAGVGTILAGSITGGGNGN---GNAATGGDAFGSFQJGGGGG----- 737
DB      693
QY      69 TITSGGNGADVQ-----QGADNSTI--ELTONGFRNNATIDQNAK--NSDITV 115
      738 ----GGYATANVAVFGTLTITGGSHAAGIVAVQSGGGGTGTASYSAGIGFTASVAV 793
DB      738
QY      116 GQYGNNA 124
      794 GGTGGNGA 802
DB      794

```

RESULT 8

AH0635

nucleation component of curlin monomers [imported] - *Salmonella enterica* subsp. *enterica*
C:Species: Salmonella enterica subsp. enterica serovar Typh
A:Note: this species has also been called Salmonella typhi
C>Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AH0635

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davies, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Garra, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A:Reference number: AB0502; PMID:21554947; PMID:11677608
A:Accession: AH0635

A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-151 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD08267.1; PID:gl6502314; GSPDB:GN00176
C:Genetics:

A:Gene: STY1180

Query Match 13.0%; Score 101.5; DB 2; Length 151;
 Best Local Similarity 30.2%; Pred. No. 0.15;

Matches 29; Conservative 17; Mismatches 49; Indels 1; Gaps 1;

QY 47 IYQGSANALALQSDARKSETTITQSGYGADVGAGADNSTIELTONGFRNNATIDQW 106

DB 49 IGVGTDNDA-RVRQSGKLISVISOEGNNRAKYDQAGNVFAYIEQTGNANDASISQS 107

QY 107 NAKNSDITVQYGGNNALVNYDQVTRVVTHEMAH 142

DB 108 AYGNSAALITQKSGNKNANTQYGTOKTAVVVKOSH 143

RESULT 9

JG6040

fimbria protein agfB precursor - *Salmonella enteritidis*
C:Species: Salmonella enteritidis
C>Date: 31-Dec-1996 #sequence_revision 31-Dec-1996 #text_change 08-Oct-1999
C:Accession: JG6040

R:Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Baner, P.A.; Kay, W.W.
J. Bacteriol. 178, 662-667, 1996
A:Title: Salmonella enteritidis agfBAC operon encoding thin, aggregative fimbriae.
A:Reference number: JG6039; PMID:96146512; PMID:8550497
A:Accession: JG6040

A:Molecule type: DNA
A:Residues: 1-151 <COU>
A:Cross-references: GB:U43280; NID:gl184712; PIDN:AAC43598.1; PID:gl184713
A:Experimental source: strain 276755-3b
C:Genetics:

A:Gene: agfB
C:Function:
A:Description: minor component of thin aggregative fimbriae
A:Note: fimbriae bind to fibronectin, plasminogen, tissue plasminogen activator
C:Keywords: fimbria
F1-21/Domain: signal sequence #status predicted <SIG>
F12-151/Product: fimbria protein agfB #status predicted <MAT>

Query Match 12.9%; Score 100.5; DB 2; Length 151;
 Best Local Similarity 30.2%; Pred. No. 0.18;

Matches 29; Conservative 17; Mismatches 49; Indels 1; Gaps 1;

QY 47 IYQGSANALALQSDARKSETTITQSGYGADVGAGADNSTIELTONGFRNNATIDQW 106

DB 49 IGVGTDNDA-RVRQSGKLISVISOEGNNRAKYDQAGNVFAYIEQTGNANDASISQS 107

QY 107 NAKNSDITVQYGGNNALVNYDQVTRVVTHEMAH 142

DB 108 AYGNSAALITQKSGNKNANTQYGTOKTAVVVKOSH 143

RESULT 10

A56038

DNA-binding protein ovo - fruit fly (*Drosophila melanogaster*)
C:Species: Drosophila melanogaster
C>Date: 01-Dec-1995 #sequence_revision 01-Dec-1995 #text_change 21-Jul-2000
C:Accession: A56038

R:Garfinkel, M.D.; Wang, J.; Liang, Y.; Mahowald, A.P.
Mol. Cell. Biol. 14, 6809-6818, 1994
A:Title: Multiple products from the shavenbaby-ovo gene region of Drosophila melanogaster
A:Reference number: A56038; PMID:95021209; PMID:7933598
A:Accession: A56038

A>Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-1028 <GAR>
A:Cross-references: GB:U11383; NID:g520526; PIDN:AAB60216.1; PID:g520527
C:Genetics:

A:Gene: ovo
A:Cross-references: FlyBase:FBgn0003028

Query Match 12.6%; Score 98.5; DB 2; Length 1028;
 Best Local Similarity 24.5%; Pred. No. 2.3;

Matches 46; Conservative 15; Mismatches 62; Indels 65; Gaps 8;

QY 3 LKVAAPFAIIVSGSALAGVFPWGGGNNHNGGNSGCPDSTLSTIYQGSANALALQSD 62

DB 59 LQNAAAAAYIMSAGS-----GGGCTGNGGAGSGPGGPGSANSGGGGGG----- 104

QY 63 ARKSETTITQSGYGADVGAGADNSTIELTONGFRNNATIDQWNAKNSDI----- 113

DB 105 -----GGNGYINGGAG-CPNNS---LDGNLNLTPASVSNYNSKFRHHHHHHOH 152

QY 114 -----TVGQ-----YGNNAAL-----VNYDQ-----LVTRVVTHEMAH 143

DB 153 NNNNNNGGQTSMMGHPFGGNPSAYGIIIXDEPDIEYDEAKIDIGTFAQNIITQATWGSS 212

QY 144 NNATANQY 151

DB 213 GQPNASAY 220

RESULT 11

S16356

ovo protein - fruit fly (*Drosophila melanogaster*)
C:Species: Drosophila melanogaster
C>Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 16-Feb-1997
C:Accession: S16356

R:Mevel-Ninio, M.; Terracol, R.; Kafatos, F.C.
EMBO J. 10, 2255-2266, 1991
A:Title: The ovo gene of Drosophila encodes a zinc finger protein required for female g
A:Reference number: S16356; PMID:91293102; PMID:1712294
A:Accession: S16356

A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-1213 <MEV>
A:Cross-references: EMBL:X59772
C:Genetics:

A:Gene: FlyBase:ovo
A:Cross-references: FlyBase:FBgn0003028
A:Introns: 931/3; 1152/3

Query Match 12.6%; Score 98.5; DB 2; Length 1213;
 Best Local Similarity 24.5%; Pred. No. 2.8;

Matches 46; Conservative 15; Mismatches 62; Indels 65; Gaps 8;

QY 3 LKVAAPFAIIVSGSALAGVFPWGGGNNHNGGNSGCPDSTLSTIYQGSANALALQSD 62

DB 422 LQNAAAAAYIMSAGS-----GGGCTGNGGAGSGPGGPGSANSGGGGGG----- 467

QY 63 ARKSETTITQSGYGADVGAGADNSTIELTONGFRNNATIDQWNAKNSDI----- 113

DB 468 -----GGNGYINGGAG-CPNNS---LDGNLNLTPASVSNYNSKFRHHHHHHOH 515

QY 114 -----TVGQ-----YGGNNAAL-----VNYDO-----LVTRVVTHERAAH 143
 Db 516 NNNNNNGGQTSMMGHPFGYGNBSAGIILKDEPDIEDAKIDIGTFAONITIOATMGSS 575
 QY 144 NNATANOY 151
 Db 576 GGRNABAY 583

RESULT 12

ADJ143

conserved hypothetical protein Atu4768 [imported] - Agrobacterium tumefaciens (strain C58)
 C/Species: Agrobacterium tumefaciens
 C/Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
 C/Accession: ADJ143

R/Mod: D.W.; Secubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.
 erge, G.; Giller, W.; Grant, C.; Guentherer, D.; Kutyavin, T.; Levy, R.; Li, M.; Mclellan
 i Karp, P.; Romero, P.; Zhang, S.
 Science 294, 2317-2323, 2001
 A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
 ster, E.W.
 A/Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
 A/Reference number: AB2577; MUID:21608550; PMID:11743193

A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-145 <KUR>
 A/Cross-references: GB:AE008689; PIDN:AL45562.1; PID:gl7743277; GSPDB:GN00187
 A/Experimental source: strain C58 (Dupont)
 C/Genetics:
 A/Gene: Atu4768

A/Map position: linear chromosome

Query Match 12.3%; Score 96; DB 2; Length 145;
 Best Local Similarity 27.1%; Pred. No. 0.41;
 Matches 38; Conservative 23; Mismatches 59; Indels 20; Gaps 5;

QY 3 LKVAFAALVSGSALAGVP-----QMG-----GGGNHNGGN-----SSGPDST 44
 Db 1 MIRKSFASALVALVGLSAAAPAMANDVRIEQYGMSSAGAGQEGYGNRIRTYONGGYNR 60
 QY 45 LSTVYGSANALALQSDARKSETTTTQSGYGGADVGQADNSTTELTONGFRNATID 104
 Db 61 IVGHQYGRNHL-S-AVQEGHDNYGTTONGNRVAVAGIGPGSNHTTLLTODGNGNIAAGV 119
 QY 105 QMNAKNSDITVGOYGNNAA 124
 Db 120 Q-VGRGCSANVSQSGNDNVA 138

RESULT 13

H98144

hypothetical protein AGR_L_228 [imported] - Agrobacterium tumefaciens (strain C58, Cerec
 C/Species: Agrobacterium tumefaciens
 C/Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
 C/Accession: H98144

R/Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Ounollo, B.; Goldman,
 A.; Liu, F.; Moliam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.;
 Science 294, 2323-2328, 2001
 A/Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
 A/Reference number: A97359; MUID:21608551; PMID:11743194
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-145 <KUR>
 A/Cross-references: GB:AE007870; PIDN:AAK8682.1; PID:gl5158413; GSPDB:GN00170
 C/Genetics:
 A/Gene: AGR_L_228

A/Map position: linear chromosome

Query Match 12.3%; Score 96; DB 2; Length 145;
 Best Local Similarity 27.1%; Pred. No. 0.41;
 Matches 38; Conservative 23; Mismatches 59; Indels 20; Gaps 5;

QY 3 LKVAFAALVSGSALAGVP-----QMG-----GGGNHNGGN-----SSGPDST 44
 Db 1 MIRKSFASALVALVGLSAAAPAMANDVRIEQYGMSSAGAGQEGYGNRIRTYONGGYNR 60

QY 45 LSTVYGSANALALQSDARKSETTTTQSGYGGADVGQADNSTTELTONGFRNATID 104
 Db 61 IVGHQYGRNHL-S-AVQEGHDNYGTTONGNRVAVAGIGPGSNHTTLLTODGNGNIAAGV 119

QY 105 QMNAKNSDITVGOYGNNAA 124
 Db 120 Q-VGRGCSANVSQSGNDNVA 138

RESULT 14

A70762

probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
 C/Species: Mycobacterium tuberculosis
 C/Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 15-Sep-2003
 C/Accession: A70762

R/Cole, S.T.; Broesch, R.; Parhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon,
 i Connor, R.; Davies, R.; Devlin, K.; Fellwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
 Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A/Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A/Reference number: A70500; MUID:98295987; PMID:9634230

A/Accession: A70762

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-678 <COL>

A/Cross-references: GB:Z74020; GB:AL123456; NID:g3261584; PIDN:CA98335.1; PID:e1300075

A/Experimental source: strain H37RV

C/Genetics:
 A/Gene: PPE

Query Match 12.1%; Score 94.5; DB 2; Length 678;
 Best Local Similarity 25.7%; Pred. No. 3.1;
 Matches 39; Conservative 24; Mismatches 48; Indels 41; Gaps 8;

QY 15 SGSAAGVYPQMGCGNHNHNGGNSGPDSTLSTVYGSANALALQ--SDARKSETTTTQ 72
 Db 384 SGSGNLG-----FNSNGNGNIGFNSG--NNNIGNGSNGGVALSVFSGSAAERS----- 432
 QY 73 SGYGGADVGQADNS-----TTELTONGFRNATIDQ--MNAKNSDITVGOYGNN 122
 Db 433 SGFGNGELSTGIGNSGQSTGWFNSATSTGWFNSGCTTNTGWFNSGCTTNTGIGNGGN- 491

QY 123 AALVNYDQLVTRVTHEM-----AHANNATAN 149
 Db 492 -----LVTSMGGLFNSGHTNTGSFN 511

RESULT 15

AH0038

probable exported protein YPO0309 [imported] - Yersinia pestis (strain CO92)
 C/Species: Yersinia pestis
 C/Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
 C/Accession: AH0038

R/Parhill, J.; Wren, B.W.; Thomson, N.R.; Tibball, R.W.; Holden, M.T.G.; Prentice, M.B
 deno-Tarrega, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
 il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
 Nature 413, 523-527, 2001
 A/Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A/Reference number: AB0001; MUID:21470413; PMID:11586360

A/Accession: AH0038

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-1238 <KUR>

A/Cross-references: GB:AL590842; PIDN:CAC89171.1; PID:gl5978410; GSPDB:GN00175

C/Genetics:
 A/Gene: YPO0309

Query Match 12.0%; Score 93.5; DB 2; Length 1238;
 Best Local Similarity 25.1%; Pred. No. 7.3;
 Matches 45; Conservative 19; Mismatches 48; Indels 67; Gaps 9;

```

QY 7 AAFPAIVVSGSALAGVVPQWGGG---NHNGGN-----SSG-----P 41
Db 409 SAFSAITATGHLTA---EWGQGMLOTHSLGDATTLIHFNIDITAMSSGISLINEANG 464
QY 42 DSTLSIYVGSANA-----LALQSDARKSETTITTOGCGN 77
Db 465 TSTADITVTGQINVSHEGITTALTTDGKTLVNDVNNIASEYDAIRLYNYNDNYAT 524
QY 78 GADVGGGADN--STIEP-----TNGFERNATIDOWNAKNSDITVG--OYGNNAL 125
Db 525 GVDGDTGADNGSTIIDLITRGALVSOQGYGINI---ETNTADITYTVGGIVHGNGTAT 580
    
```

Search completed: March 11, 2004, 18:42:06
 Job time : 11.3 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:14:48 ; Search time 6.3 Seconds

(without alignments)
1248.031 million cell updates/sec

Title: US-09-543-407-12

Perfect score: 779

Sequence: 1 MKLLKVAAPAAIVSGSALA.....VTRVTHMAHANNATANOY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	695	89.2	151	1	CSGA_SALTY
2	520	66.8	151	1	CSGA_SALTY
3	501.5	64.4	152	1	CSGA_SALTY
4	101.5	13.0	151	1	CSGA_SALTY
5	100.5	12.9	151	1	CSGA_SALTY
6	98.5	12.6	151	1	CSGA_SALTY
7	97	12.5	156	1	CSGA_SALTY
8	94.5	12.1	156	1	CSGA_SALTY
9	92	11.8	156	1	CSGA_SALTY
10	91.5	11.7	151	1	CSGA_SALTY
11	89.5	11.5	151	1	CSGA_SALTY
12	88	11.4	151	1	CSGA_SALTY
13	86	11.3	151	1	CSGA_SALTY
14	87	11.2	151	1	CSGA_SALTY
15	86	11.0	151	1	CSGA_SALTY
16	85	10.9	151	1	CSGA_SALTY
17	84.5	10.8	151	1	CSGA_SALTY
18	84	10.8	151	1	CSGA_SALTY
19	84	10.8	151	1	CSGA_SALTY
20	83.5	10.7	151	1	CSGA_SALTY
21	83	10.7	151	1	CSGA_SALTY
22	83	10.7	151	1	CSGA_SALTY
23	83	10.7	151	1	CSGA_SALTY
24	83	10.7	151	1	CSGA_SALTY
25	82.5	10.6	151	1	CSGA_SALTY
26	82.5	10.6	151	1	CSGA_SALTY
27	82	10.5	151	1	CSGA_SALTY
28	81.5	10.5	151	1	CSGA_SALTY
29	81.5	10.5	151	1	CSGA_SALTY
30	81	10.4	151	1	CSGA_SALTY
31	81	10.4	151	1	CSGA_SALTY
32	81	10.4	151	1	CSGA_SALTY
33	81	10.4	151	1	CSGA_SALTY

34	81	10.4	1460	1	PMPC_CHLMTU
35	80.5	10.3	467	1	Y442_MYCTU
36	80.5	10.3	2003	1	YDBA_ECOLI
37	80	10.3	362	1	P35_MYCPE
38	80	10.3	594	1	SUH_DROME
39	80	10.3	955	1	PRU_DROME
40	80	10.3	1778	1	N185_SCHPO
41	79.5	10.2	485	1	Y136_TREPA
42	79.5	10.2	760	1	YBIL_ECOLI
43	79	10.1	572	1	FLAB_CAMCO
44	79	10.1	933	1	NPA3_HUMAN
45	79	10.1	1113	1	N116_YEAST

ALIGNMENTS

RESULT 1	CSGA_SALTY	STANDARD	PRT	151 AA.
AC	P55225			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Major curlin subunit precursor (Fimbrin SEF17).			
GN	CSGA OR AGFA OR STM1144 OR STY1181 OR T1776.			
OS	Salmonella typhimurium,			
OS	Salmonella typhi, and			
OS	Salmonella enteritidis.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OC	Enterobacteriaceae; Salmonella.			
OX	NCBI_TaxID=602, 601, 592;			
LN	[1]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=S. typhimurium; STRAIN=SR-11;			
RX	MEDLINE=98117058; PubMed=9457880;			
RA	Romling U., Bian Z., Hammar M., Sierralta W.D., Normark S.;			
RT	"Curli fibers are highly conserved between Salmonella typhimurium and			
RT	Escherichia coli with respect to operon structure and regulation.";			
RL	J. Bacteriol. 180:722-731(1998).			
LN	[2]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=S. typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;			
RX	MEDLINE=21534948; PubMed=11677609;			
RA	McClainland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,			
RA	Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,			
RA	Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,			
RA	Ryan B., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,			
RA	Waterston R., Wilson R.K.;			
RT	"Complete genome sequence of Salmonella enterica serovar Typhimurium			
RT	LT2.";			
RL	Nature 413:852-856(2001).			
LN	[3]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=S. typhi; STRAIN=CT18;			
RX	MEDLINE=21534947; PubMed=11677608;			
RA	Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,			
RA	Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sabañia M.,			
RA	Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,			
RA	Cronin A., Davis P., Davies K.M., Dowd L., White N., Farrar J.,			
RA	Fellwell T., Hamlin N., Haque A., Hien T.T., Holtroyd S., Jagels K.,			
RA	Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,			
RA	Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,			
RA	Whithead S., Barrall B.G.;			
RT	"Complete genome sequence of a multiple drug resistant Salmonella			
RT	enterica serovar Typhi CT18.";			
RL	Nature 413:848-852(2001).			
LN	[4]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=S. typhi; STRAIN=Ty2 / ATCC 700931;			
RX	MEDLINE=22531167; PubMed=12644504;			
RA	Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,			
RA	Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;			

RT "Comparative genomics of *Salmonella enterica* serovar Typhi strains Ty2
 RT and Cti8".
 RL J. Bacteriol. 185:2330-2337(2003).
 RN (5)
 RP SEQUENCE FROM N.A.
 RC SPECIES=S. enteritidis; STRAIN=27655-3B;
 RX MEDLINE=96146512; PubMed=8550497;
 RA Collinson S.K., Clouthier S.C., Doran J.L., Baner P.A., Kay W.W.;
 RT "Salmonella enteritidis agfBAC operon encoding thin, aggregative
 RT fimbriae".
 RL J. Bacteriol. 178:662-667(1996).
 RN (6)
 RP SEQUENCE OF 21-151 FROM N.A.
 RC SPECIES=S. enteritidis; STRAIN=27655-3B;
 RX MEDLINE=94013373; PubMed=8104955;
 RA Doran J.L., Collinson S.K., Burian J., Sarlos G., Todd E.C.D.,
 RA Munro C.K., Kay C.W., Baner P.A., Peterkin P.I., Kay W.W.;
 RT "DNA-based diagnostic tests for *Salmonella* species targeting agfA,
 RT the structural gene for thin, aggregative fimbriae".
 RL J. Clin. Microbiol. 31:2263-2273(1993).
 RN (7)
 RP SEQUENCE OF 21-33.
 RC SPECIES=S. enteritidis; STRAIN=27655-3B;
 RX MEDLINE=91310586; PubMed=1677357;
 RA Collinson S.K., Emsedy L., Mueller K.M., Trust T.J., Kay W.W.;
 RT "Purification and characterization of thin, aggregative fimbriae from
 RT *Salmonella enteritidis*".
 RL J. Bacteriol. 173:4773-4781(1991).
 CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 CC FIBRONECTIN.
 CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: AJ002301; GAA05317.1; -
 DR EMBL: AE008749; AAL20074.1; -
 DR EMBL: AL627269; CAD08268.1; -
 DR EMBL: AE016840; AAO69399.1; -
 DR EMBL: U43280; AAC43599.1; -
 DR PIR: JG6039; JG6039.
 DR StyGene; SG10608; csGA.
 KW Fimbria; Signal; Complete proteome.
 FT SIGNAL 1 20
 FT CHAIN 1 151
 FT CONFLICT 134 151 SVMVROVGFNNATANQY -> DSYTVQVAS (IN
 FT REF. 6).
 SQ SEQUENCE 151 AA; 15305 MW; B7DACOD16B621359 CMC64;

Query Match 89.2%; Score 695; DB 1; Length 151;
 Best Local Similarity 90.1%; Fred. No. 4.6e-51;
 Matches 136; Conservative 4; Mismatches 11; Indels 0; Gaps 0;

QY 1 MKLKVAAFAAIVVSGSALAGVPOWGGGNGHNGSGSDPSTLSTIYOGSANAALALQ 60
 DB 1 MKLKVAAFAAIVVSGSALAGVPOWGGGNGHNGSGSDPSTLSTIYOGSANAALALQ 60
 QY 61 SDARKSETTITQSGYGAGADVGGADNSTLELTONGFRNNATTIDOWNAKNSDITVQYGG 120
 DB 61 SDARKSETTITQSGYGAGADVGGADNSTLELTONGFRNNATTIDOWNAKNSDITVQYGG 120
 QY 121 NNAALVNVYDQLVTRVYVTHEMAHANNATANQY 151
 DB 121 NNAALVNVYDQLVTRVYVTHEMAHANNATANQY 151

RESULT 2
 ID CSGA_ECOLI STANDARD; PRT; 151 AA.
 AC P28307;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Major curlin subunit precursor.
 GN CSGA OR B1042.
 OS *Escherichia coli*.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; *Escherichia*.
 OX NCBI_TaxID=562;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / W3110;
 RX MEDLINE=93211294; PubMed=8459772;
 RA Olsen A., Arngvist A.;
 RT "The Rpos sigma factor relieves H-NS-mediated transcriptional
 RT repression of csGA, the subunit gene of fibronectin-binding curli in
 RT *Escherichia coli*".
 RL Mol. Microbiol. 7:523-536(1993).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MC4100;
 RX MEDLINE=96414468; PubMed=8817489;
 RA Hammar M., Arngvist A., Bian Z., Olsen A., Normark S.;
 RT "Expression of two csG operons is required for production of
 RT fibronectin- and Congo red-binding curli polymers in *Escherichia coli*
 RT K-12".
 RL Mol. Microbiol. 18:661-670(1995).
 RN (3)
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of *Escherichia coli* K-12";
 RL Science 277:1234-1238(1997).
 RN (4)
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97061202; PubMed=8905232;
 RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
 RA Ikemoto K., Inada T., Itoh T., Kajinaka M., Kanai K., Kashimoto K.,
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
 RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horiiuchi T.;
 RT "A 718-kb DNA sequence of the *Escherichia coli* K-12 genome
 RT corresponding to the 12.7-28.0 min region on the linkage map";
 RL DNA Res. 3:1137-1155(1996).
 RN (5)
 RP SEQUENCE OF 21-40.
 RC STRAIN=K12 / YMEJ;
 RX MEDLINE=93023873; PubMed=1357528;
 RA Arngvist A., Olsen A., Pfeiffer J., Russell D.G., Normark S.;
 RT "The Crl protein activates cryptic genes for curli formation and
 RT fibronectin binding in *Escherichia coli* HB101".
 RL Mol. Microbiol. 6:2443-2452(1992).
 RN (6)
 RP SEQUENCE OF 21-31.
 RC MEDLINE=91310586; PubMed=1677357;
 RA Collinson S.K., Emsedy L., Trust T.J., Kay W.W.;
 RT "Purification and characterization of thin, aggregative fimbriae from
 RT *Salmonella enteritidis*".
 RL J. Bacteriol. 173:4773-4781(1991).
 CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 CC FIBRONECTIN.

```

CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: I04979; AAA32616.1; -
DR EMBL: X90754; CAA62282.1; -
DR EMBL: AE000205; AAC74126.1; -
DR EMBL: D90741; BAA35832.1; -
DR EMBL: D90742; BAA35840.1; -
DR PIR: S70788; S70788.
DR EcGene: EG11489; csga.
KM Rimbria; Signal; Complete proteome.
FT SIGNAL 1 20
FT CHAIN 21 151 MAJOR CURLIN SUBUNIT.
FT CONFICT 7 7 A -> B (IN REF. 1).
SQ SEQUENCE 151 AA; 15049 MW; C003470D208D395F CRC64;

Query Match 66.8%; Score 520; DB 1; Length 151;
Best Local Similarity 66.8%; Pred. No. 1.5e-36;
Matches 101; Conservative 21; Mismatches 29; Indels 0; Gaps 0;

QY 1 MKLLKVAAPFAIVVSGSALAGVVPOMGGGNGHNGSGPDSITSIYOGSANAALALQ 60
DB 1 MKLLKVAAPFAIVVSGSALAGVVPYGGGNGHGGGNGSGPSELSLTYOGGGSALALQ 60
QY 61 SDARKSETTITQSGYGNGADVGGADNSTIELTQNGFRNNATIDQNNAKNSDITVGGYG 120
DB 61 TDARNSDLITTHGGGNGADVGGSDSSIDLTRGSGNSATLDDQNGKXSEMTVVGFG 120

QY 121 NNALVNTDQLVTRVVTTHMAHANNATANOY 151
DB 121 GNGAAYDQTASNSVTNVTVGFGNNATANOY 151

RESULT 3
CSGA_ECO57 STANDARD; PRT; 152 AA.
AC 093024;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Major curlin subunit precursor.
CSGA OR Z1676 OR ECS1420.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
NCBI_TaxID=83334;
RX MEDLINE=21074935; PubMed=11206551;
RX STRAIN=0157:H7 / ATCC 700927;
RX Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RX Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RX Posfai G., Hackert J., Klink S., Boutin A., Shao Y., Miller L.,
RX Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouis K.,
RX Apodoca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RX Welch R.A., Blattner F.R., Linn J., Schwartz D.C.,
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.",
RL Nature 409:529-533(2001).

```

```

RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=0157:H7 / RMD 050952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,
RA Kihara S., Shiba T., Hattori M., Shingawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.",
RL DNA Res. 8:11-22(2001)

CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN.
CC -----
CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF275733; AAK53212.1; -
DR EMBL: AE005315; AAG55788.1; -
DR EMBL: AP002554; BAB34843.1; -
DR PIR: D90806; D90806.
DR PIR: H85665; H85665.
KW Rimbria; Signal; Complete proteome.
FT SIGNAL 1 20 BY SIMILARITY.
FT CHAIN 21 152 MAJOR CURLIN SUBUNIT.
SQ SEQUENCE 152 AA; 15099 MW; EE2D2D94DDE91243 CRC64;

Query Match 64.4%; Score 501.5; DB 1; Length 152;
Best Local Similarity 65.8%; Pred. No. 5.1e-35;
Matches 100; Conservative 21; Mismatches 30; Indels 1; Gaps 1;

QY 1 MKLLKVAAPFAIVVSGSALAGVVPOMGGGNGHNGSGPDSITSIYOGSANAALALQ 59
DB 1 MKLLKVAAPFAIVVSGSALAGVVPYGGGNGHGGGNGSGPSELSLTYOGGGSALALQ 60
QY 60 QSDARKSETTITQSGYGNGADVGGADNSTIELTQNGFRNNATIDQNNAKNSDITVGGYG 119
DB 61 QADARNSDLITTHGGGNGADVGGSDSSIDLTRGSGNSATLDDQNGKXSEMTVVGFG 120

QY 120 NNALVNTDQLVTRVVTTHMAHANNATANOY 151
DB 121 GNGAAYDQTASNSVTNVTVGFGNNATANOY 152

RESULT 4
CSGB_SALTI STANDARD; PRT; 151 AA.
AC 0827M3;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Minor curlin subunit precursor.
CSGB OR STY1180 OR T1777.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
NCBI_TaxID=601;
RX MEDLINE=21534947; PubMed=11677608;
RX STRAIN=CT18;
RX Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RX Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebatina M.,
RX Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RX Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,

```

RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
 RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Pary C.,
 RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
 RA Whitehead S., Barrell B.G.;
 RT "Complete genome sequence of a multiple drug resistant *Salmonella*
 RT enterica serovar Typhimurium CT18.";
 RL Nature 413:848-852(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Ty2 / ATCC 700931;
 RX MEDLINE=22531367; PubMed=12644504;
 RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
 RA Burland V., Kodyanski V., Schwartz D.C., Blattner F.R.;
 RT "Comparative genomics of *Salmonella enterica* serovar Typhimurium Ty2
 RT and CT18.";
 RL J. Bacteriol. 185:2330-2337(2003).
 CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
 CC CURLIN MONOMERS.
 CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AL627269; CAD08267.1; -;
 DR EMBL; AB016840; AAO69400.1; -;
 DR Fimbrin; Signal; Complete proteome.
 FT SIGNAL 1 21
 FT CHAIN 1 21
 FT SEQUENCE 151 AA; 16254 MW; 161C54326B573495 CRC64;
 SQ
 Query Match 13.0%; Score 101.5; DB 1; Length 151;
 Best Local Similarity 30.2%; Pred. No. 0.071; Mismatches 49; Indels 1; Gaps 1;
 Matches 29; Conservative 17; Mismatches 49; Indels 1; Gaps 1;
 QY 47 IYQYGSANALALQSDARKSETTITQSGYNGADVGAGDAGNSTIELTNGFNNAITDQW 106
 DB 49 IGVGTDNDA-RVRQSGSKLTVISQEGNNRAKVDQAGNVPAYIEGTGANNASISQS 107
 QY 107 NAKNSDITVGQYGNNAALVNYDQLVTRVYTHEMAH 142
 DB 108 AYGNSPAIIQKSGNKANITQYGTQKTAVVVQKQSH 143
 RESULT 5
 CSGB_SALTY STANDARD; PRT; 151 AA.
 ID _CSGB_SALTY STANDARD; PRT; 151 AA.
 AC P55226; 1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Minor curlin subunit precursor (Fimbrin SEF17 minor subunit).
 GN CSGB OR AGFB OR STM1143.
 OS *Salmonella typhimurium*, and
 OS *Salmonella enteritidis*.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; *Salmonella*.
 OC NCBI_TaxID=602, 592;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S.typhimurium; STRAIN=SR-11;
 RX MEDLINE=96117058; PubMed=9457880;
 RA Romling U., Bian Z., Hammar M., Sierraalta W.D., Normark S.;
 RT "Curli fibers are highly conserved between *Salmonella typhimurium* and
 RT *Escherichia coli* with respect to operon structure and regulation.";
 RL J. Bacteriol. 180:722-731(1998).

EN [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S.typhimurium; STRAIN=L72 / SGSC1412 / ATCC 700720;
 RX MEDLINE=21531948; PubMed=11677609;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Portolillo S., Ali J., Dante M., Du F., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
 RA Ryan B., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RT "Complete genome sequence of *Salmonella enterica* serovar Typhimurium
 RT L72.";
 RL Nature 413:852-856(2001).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S.enteritidis; STRAIN=27655-3B;
 RX MEDLINE=96146512; PubMed=8550497;
 RA Collinson S.K., Clouthier S.C., Doran J.L., Banerjee P.A., Kay M.W.;
 RT "Salmonella enteritidis agfBAC operon encoding thin, aggregative
 RT fimbriae.";
 RL J. Bacteriol. 178:662-667(1996).
 CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
 CC CURLIN MONOMERS.
 CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AJ002301; CA05316.1; -;
 DR EMBL; AE008749; AAL20073.1; -;
 DR EMBL; U43280; AAC43598.1; -;
 DR PIR; JC6040; JC6040.
 DR ScyGene; SG10609; CSGB.
 DR Fimbrin; Signal; Complete proteome.
 FT SIGNAL 1 21
 FT CHAIN 1 21
 FT SEQUENCE 151 AA; 16182 MW; C0FC5430B6DD361D CRC64;
 SQ
 Query Match 12.9%; Score 100.5; DB 1; Length 151;
 Best Local Similarity 30.2%; Pred. No. 0.086; Mismatches 49; Indels 1; Gaps 1;
 Matches 29; Conservative 17; Mismatches 49; Indels 1; Gaps 1;
 QY 47 IYQYGSANALALQSDARKSETTITQSGYNGADVGAGDAGNSTIELTNGFNNAITDQW 106
 DB 49 IGVGTDNDA-RVRQSGSKLTVISQEGNNRAKVDQAGNVPAYIEGTGANNASISQS 107
 QY 107 NAKNSDITVGQYGNNAALVNYDQLVTRVYTHEMAH 142
 DB 108 AYGNSPAIIQKSGNKANITQYGTQKTAVVVQKQSH 143
 RESULT 6
 OVO_DROME STANDARD; PRT; 1028 AA.
 ID _OVO_DROME STANDARD; PRT; 1028 AA.
 AC P51521; Q9XZU4;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE OVO protein (shaven baby protein).
 GN OVO OR SVB.
 OS *Drosophila melanogaster* (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Preygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; *Drosophila*.
 OC NCBI_TaxID=7227;
 RN [1]

RP SEQUENCE FROM N.A.
 RC TISSUE=Ovary;
 RX MEDLINE=95021209; PubMed=7935398;
 RA Garfinkel M.D., Wang J., Liang Y., Mahowald A.P.;
 RT "Multiple products from the shavenbaby-ovo gene region of Drosophila
 melanogaster: relationship to genetic complexity.";
 RN Mol. Cell. Biol. 14:6809-6818 (1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Oregon-R;
 RX MEDLINE=91293102; PubMed=1712294;
 RA Meyer-Ninio M.T.M., Terracol R., Katatos F.C.;
 RT "The ovo gene of Drosophila encodes a zinc finger protein required
 for female germ line development.";
 RL EMBO J. 10:2259-2266 (1991).
 CC -1- FUNCTION: REQUIRED FOR SURVIVAL AND DIFFERENTIATION OF FEMALE GERM
 CC LINE CELLS. PLAYS A ROLE IN GERM LINE SEX DETERMINATION.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
 CC -1- DEVELOPMENTAL STAGE: FIRST APPEARS IN THE GERMARIIUM AND
 CC ACCUMULATES IN NURSE CELLS DURING OOGENESIS. STORED IN THE EGG,
 CC BUT IS RAPIDLY LOST IN THE EMBRYOS EXCEPT FOR ITS CONTINUED
 CC PRESENCE IN THE GERM LINE PRECURSOR POLE CELLS.
 CC -1- SIMILARITY: Contains 4 C2H2-type zinc fingers.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL, U11383; CAB60216.1; -;
 DR EMBL, X59772; CAB36921.1; ALT_SEQ.
 DR PIR: A56038; A56038.
 DR HSSP: P07248; 2ADR.
 DR TRANSFAC: T00669; -;
 DR FlyBase: FBgn003028; ovo.
 DR InterPro: IPR007087; Znf_C2H2.
 DR Pfam: PF00096; zf-C2H2; 3.
 DR SMART: SM00355; ZNF_C2H2; 4.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 3.
 DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 3.
 DR KMWinc-finger; Metal-binding; DNA-binding; Repeat; Nuclear protein;
 KM Transcription regulation.
 FT DOMAIN 62 65 POLY-ALA.
 FT DOMAIN 72 77 POLY-GLY.
 FT DOMAIN 80 85 POLY-GLY.
 FT DOMAIN 98 108 POLY-GLY.
 FT DOMAIN 144 152 POLY-HIS.
 FT DOMAIN 153 159 POLY-ASN.
 FT DOMAIN 336 339 POLY-GLN.
 FT DOMAIN 347 353 POLY-GLN.
 FT DOMAIN 357 361 POLY-GLN.
 FT DOMAIN 410 414 POLY-GLN.
 FT DOMAIN 418 422 POLY-GLN.
 FT DOMAIN 426 432 POLY-GLN.
 FT DOMAIN 445 453 POLY-GLN.
 FT DOMAIN 456 459 POLY-GLN.
 FT DOMAIN 466 474 POLY-GLN.
 FT DOMAIN 497 517 POLY-ALA.
 FT DOMAIN 524 529 POLY-SER.
 FT DOMAIN 549 558 POLY-ALA.
 FT DOMAIN 639 651 POLY-ALA.
 FT DOMAIN 717 725 POLY-ALA.
 FT DOMAIN 797 802 POLY-GLN.
 FT DOMAIN 820 823 POLY-GLN.
 FT DOMAIN 826 832 POLY-GLN.
 FT ZN_FING 874 896 C2H2-TYPE 1.
 FT ZN_FING 902 924 C2H2-TYPE 2.
 FT ZN_FING 930 953 C2H2-TYPE 3.
 FT ZN_FING 969 992 C2H2-TYPE 4.
 FT CONFLICT 647 647 A -> R (IN REF. 2).

SQ SEQUENCE 1028 AA; 110620 MW; D7068BB2BC0F6F77 CRC64;
 Query Match 12.6%; Score 98.5; DB 1; Length 1028;
 Best Local Similarity 24.5%; Pred. No. 1.1;
 Matches 46; Conservative 15; Mismatches 62; Indels 65; Gaps 8;
 QY 3 LKVAAPAAIVVSGSALAGVPMQGGGNNHGGSSGPDSTLSTIYQGSANALALQSD 62
 DB 59 LQVAAAAYIMSAGS-----GGCTGNGGCGAGSGPGSPANSGGGGGG----- 104
 QY 63 ARKSETTIOSGYGADYGGADNSTITLTQGFNNATIIDMANKNSI----- 113
 DB 105 -----GGNGYINGGVG-GPNN-----LDGNNLINFASVSNYSNSKFEHHHHHHQ 152
 QY 114 -----TVGG-----YGNNAL-----VNYDQ-----LVTVVTHEMAHA 143
 DB 153 NNNNNNGGOTSMGHPFGGNSAIGIILKDEPDLEYPDADIDGTFAQNTIQTMGSS 212
 QY 144 NNATANQY 151
 DB 213 GGFNASAY 220
 RESULT 7
 OMPB_RICUA STANDARD; PRT; 1656 AA.
 ID OMPB_RICUA
 AC 006653;
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Outer membrane protein B precursor (168 kDa surface-layer protein)
 DE (Surface protein antigen) (Cell surface antigen 5) (Scs5) (rOmpB)
 DE (rOmp B) [Contains: 120 kDa surface-exposed protein (Surface protein
 DE antigen) (120 kDa outer membrane protein ompb); 32 kDa beta peptidol].
 GN OMPB.
 OS Rickettsia japonica.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 OX NCBI_TaxID=35790;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TH;
 RA Uchiyama T.;
 RT "Sequencing of the gene encoding the protein rOmp B of Rickettsia
 RT japonica.";
 RT Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: THE 120 kDa SURFACE-EXPOSED PROTEIN IS A MAJOR
 CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
 CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (BY
 CC similarity).
 CC -1- FUNCTION: THE 32 kDa BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR
 CC (BY similarity).
 CC -1- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-
 CC layer with hexagonal symmetry.
 CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAL OMPA/OMPB FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL, AB003681; BAA20138.1; -;
 DR InterPro: IPR006315; Autotransport.
 DR InterPro: IPR005546; Autotransporter.
 DR Pfam: PF03797; Autotransporter; 1.
 DR TIGRfams: TIGR01414; autotrans_bar1; 2.
 KM Antigen; S-layer; Cell wall.
 FT CHAIN 1 1338 120 kDa SURFACE-EXPOSED PROTEIN.
 FT CHAIN 1339 1656 32 kDa BETA PEPTIDE.
 FT DOMAIN 528 533 POLY-GLY.

SQ SEQUENCE 1656 AA; 168097 MW; 3132A69C9DD5999F CRC64;
 Query Match 12.5%; Score 97; DB 1; Length 1656;
 Best Local Similarity 28.3%; Pred. No. 2.6;
 Matches 43; Conservative 19; Mismatches 52; Indels 38; Gaps 7;
 QY 6 VAAFAIVSGSALAGVPPWGGGNNHNGSGSDSTLSIYQVSANALALQSDARK 65
 DB 509 VLAAGATLTDGSAIT-----TGDJGNGGG-----GALAQSTTLNDARK 547
 QY 66 SETTTTQSG-----YGCADVGGGADNSSTLTETONGFRNNATIDOMWAKNSDITVG--OYG 119
 DB 548 ---TLTLGANNITISANGGTINFPANGTITKLTST--QNNITVD-----CDLAIATDQTG 596
 QY 120 GNNAAALVNDQVLTTRYVTHEMAANATANQY 151
 DB 597 VVDASSLTNAQTLLTISGTITGIGANNNTTLGQF 628
 RESULT 8
 YF48 MYCTU STANDARD; PRT; 678 AA.
 AC 010776;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 15-UTL-1999 (Rel. 38, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Hypothetical PPE-family protein RV1548c/MT1599.
 GN RV1548C OR MT1599 OR MTCY48.17.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37Rv;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Bisham D., Brown D., Chillingworth T., Connor R.,
 RA Badcock K., Basham D., Barron P., Barry C.E. III, Tekka R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornby T., Jagels K., Krogh A., McLellan J., Moule S., Murphy L.,
 RA Oliver S., Osborne U., Quail M.A., Rajandream W.A., Rogers J.,
 RA Ruster S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sutcliffe J.B., Taylor K., Whitehead S., Barrett B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 RL Nature 393:537-544 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RX MEDLINE=22206494; PubMed=12218036;
 RA Fletschmann R.D., Alland D., Bisen U.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
 RA Bishai W., Jacobs W.R. Jr., Venter J.C., Kouri H., Gill U., Mikula A.,
 RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains."
 RL J. Bacteriol. 184:5479-5490 (2002).
 RN [3]
 RP -1 SIMILARITY: Belongs to the mycobacterial PPE family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.ebi.ac.uk/announcements/](http://www.ebi.ac.uk/announcements)
 CC or send an email to license@ebi.ac.uk).
 CC EMBL; Z74020; CAA98335.1;
 DR EMBL; AE007026; AAK45866.1; ALT_INIT.
 DR PIR; A70762; A70762.
 DR TIGR; MT1599; -.

DR TubercList; RV1548c; -
 DR InterPro; IPR000030; Microbac_PPE.
 DR InterPro; IPR002989; Mycobac_Pentapep.
 DR Pfam; PF01469; Pentapeptide_2; 11.
 DR Pfam; PF00823; PPE; 1.
 KM Hypothetical protein; Transmembrane; Complete proteome.
 FT TRANSMEM 14 34
 FT TRANSMEM 180 200 POTENTIAL.
 FT CONFLICT 258 258 D -> G (IN REF. 2).
 SQ SEQUENCE 678 AA; 66736 MW; 209F1593D5253A2 CRC64;
 Query Match 12.1%; Score 94.5; DB 1; Length 678;
 Best Local Similarity 25.7%; Pred. No. 1.5;
 Matches 39; Conservative 24; Mismatches 48; Indels 41; Gaps 8;
 QY 15 SGSLAGVPPWGGGNNHNGSGSDSTLSIYQVSANALALQ--SDARKSETTTQ 72
 DB 384 SGSGNIG-----FNSNGNIGFNSG--NNNIGKNSGNGVGLSVFGSSARS----- 432
 QY 73 SGYNGADVGGGADNS-----TTLTONGFRNNATIDQ--WNAKNSDITVGQYGN 122
 DB 433 SGFGNSGELSTGIGNSGQLSTGWFNSATSTGWFNSGTTNTGWFNSGTTNTGIGNSGN- 491
 QY 123 AALVNDQVLTTRYVTHEM-----AHANNTAN 149
 DB 492 -----LVTSWGLFNSGHTNTGSFN 511
 RESULT 9
 TNK1 HUMAN STANDARD; PRT; 1327 AA.
 AC 095271; 095272;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Tankyrase 1 (EC 2.4.2.30) (TANK1) (TANKS-1) (TRF1-
 DE interacting ankyrin-related ADP-ribose polymerase).
 GN TNKS OR TNKS1 OR TINI OR TINF1 OR PARPL.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
 RC TISSUE=Testis;
 RX MEDLINE=99040105; PubMed=9822378;
 RA Smith S., Giriat L., Schmitt A., de Lange T.;
 RT "Tankyrase, a poly(ADP-ribose) polymerase at human telomeres."
 RL Science 282:1484-1487 (1998).
 RN [2]
 RP SUBCELLULAR LOCATION.
 RX MEDLINE=99454782; PubMed=10523501;
 RA Smith S., de Lange T.;
 RT "Cell cycle dependent localization of the telomeric PARP, tankyrase,
 RT to nuclear pore complexes and centrosomes."
 RL J. Cell Sci. 112:3649-3656 (1999).
 RN [3]
 RP FUNCTION, AND PHOSPHORYLATION.
 RX MEDLINE=20556282; PubMed=10988299;
 RA Chi N.-W., Lodish H.F.;
 RT "Tankyrase is a Golgi-associated mitogen-activated protein kinase
 RT substrate that interacts with IRAP in GLUT4 vesicles."
 RL J. Biol. Chem. 275:38437-38444 (2000).
 RN [4]
 RP FUNCTION, AND MUTAGENESIS OF HIS-1194 AND GLU-1291.
 RX MEDLINE=21602874; PubMed=11739745;
 RA Cook B.D., Dynek J.N., Chang W., Shostak G., Smith S.;
 RT "Role for the related poly(ADP-Ribose) polymerases tankyrase 1 and 2
 RT at human telomeres."
 RL Mol. Cell. Biol. 22:332-342 (2002).
 CC -1- FUNCTION: May regulate vesicle trafficking and modulate the
 CC subcellular distribution of SLC24A4/GLUT4-vesicles. Has PARP
 CC activity and can modify TRF1, and thereby contribute to the

CC regulation of telomere length.
 CC -1- CATALYTIC ACTIVITY: NAD(+) + {ADP-D-riboseyl}(N)-acceptor =
 CC nicotinamide + {ADP-D-riboseyl}(N+1)-acceptor.
 CC -1- SUBUNIT: Oligomerizes and associates with TNKS2. Interacts with
 CC the cytoplasmic domain of LMNB1/Oterase in SLIC2A4/GLUT4-vesicles.
 CC -1- BINDS TO THE N-terminus of telomeric TRF1 via the ANK repeats.
 CC with juxtanuclear SLIC2A4/GLUT4-vesicles. A minor proportion is
 CC also found at nuclear pore complexes and around the pericentriolar
 CC matrix of mitotic centrosomes. During interphase, a small fraction
 CC of TNKS is found in the nucleus, associated with TRF1.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Name=1;
 CC isoId=O95271-1; Sequence=Displayed;
 CC Name=2;
 CC isoId=O95271-2; Sequence=VSP_004538, VSP_004539;
 CC Note=No experimental confirmation available;
 CC -1- TISSUE SPECIFICITY: Ubiquitous; highest levels in testis.
 CC -1- PTM: Upon insulin-stimulation, phosphorylated on serine residues
 CC by MAPK kinases.
 CC -1- PTM: ADP-ribosylated (-auto).
 CC -1- SIMILARITY: Belongs to the PARP family.
 CC -1- SIMILARITY: Contains 15 ANK repeats.
 CC -1- SIMILARITY: Contains 1 sterile alpha motif (SAM) domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AF082556; AAC79841.1; -;
 CC EMBL: AF082557; AAC79842.1; -;
 CC EMBL: AF082558; AAC79843.1; -;
 CC EMBL: AF082559; AAC79844.1; -;
 CC HSSP: Q00420; IAWC.
 CC Gene: HGNC:11941; TNKS.
 CC MIM: 603303; -;
 CC GO: GO:0000781; C:chromosome, telomeric region; IDA.
 CC GO: GO:0003950; F:NAD ADP-ribosyltransferase activity; IDA.
 CC GO: GO:0005515; F:protein binding; IPI.
 CC GO: GO:0007004; P:telomerase-dependent telomere maintenance; NAS.
 CC InterPro: IPR002110; ANK.
 CC InterPro: IPR001660; SAM.
 CC Pfam: PF00023; ank; 19.
 CC Pfam: PF00536; SAM; 1.
 CC PRINTS: PR01415; ANKYRIN.
 CC SMART: SM00248; ANK; 17.
 CC SMART: SM00454; SAM; 1.
 CC PROSITE: PSS0088; ANK_REPEAT; 15.
 CC PROSITE: PSS0287; ANK_REPEAT_REGION; 1.
 CC PROSITE: PSS0105; SAM_DOMAIN; 1.
 CC TRANSFERASE: Glycosyltransferase; NAD; Golgi stack; Telomere;
 CC Nuclear protein; Repeat; ANK repeat; ADP-ribosylation;
 CC Phosphorylation; Alternative splicing.
 CC FT REPEAT 215 247 ANK 1.
 CC FT REPEAT 248 280 ANK 2.
 CC FT REPEAT 281 313 ANK 3.
 CC FT REPEAT 368 400 ANK 4.
 CC FT REPEAT 401 433 ANK 5.
 CC FT REPEAT 434 466 ANK 6.
 CC FT REPEAT 467 499 ANK 7.
 CC FT REPEAT 521 556 ANK 8.
 CC FT REPEAT 557 589 ANK 9.
 CC FT REPEAT 590 622 ANK 10.
 CC FT REPEAT 683 715 ANK 11.
 CC FT REPEAT 716 748 ANK 12.
 CC FT REPEAT 749 781 ANK 13.
 CC FT REPEAT 836 868 ANK 14.
 CC FT REPEAT 869 901 ANK 15.
 CC FT REPEAT 902 934 ANK 15.
 CC RT

FT DOMAIN 1030 1089 SAM.
 FT DOMAIN 1176 1327 PARP.
 FT DOMAIN 9 14 POLY-HIS.
 FT DOMAIN 27 34 POLY-PRO.
 FT DOMAIN 128 134 POLY-SER.
 FT DOMAIN 137 145 POLY-SER.
 FT VARSPPLIC 641 643 EST -> GHS (in isoform 2).
 FT VARSPPLIC 644 1327 /FTID=VSP_004538.
 FT VARSPPLIC 644 1327 Missing (in isoform 2).
 FT MOTAGEN 1184 1184 /FTID=VSP_004539.
 FT MOTAGEN 1291 1291 H->A: LOSS OF ACTIVITY; WHEN ASSOCIATED
 FT MOTAGEN 1291 1291 WITH A-1291.
 FT MOTAGEN 1291 1291 E->A: LOSS OF ACTIVITY; WHEN ASSOCIATED
 FT MOTAGEN 1291 1291 WITH A-1184.
 SQ SEQUENCE 1327 AA; 142010 MW; E14DE985C710B957 CRC64;
 FT
 Query Match 11.8%; Score 92; DB 1; Length 1327;
 Best Local Similarity 30.4%; Pred. No. 5.2; Mismatches 8; Gaps 4;
 Matches 35; Conservative 15; Indels 8; Gaps 4;
 QY 6 VAAFAAI-VVSGSALAGVVPQWGGGNGGNSGSPDSTLSIYQGSANALALQSDAR 64
 DB 99 VAAAPVPAVSTSSAGVAPNPAGSGSNSSPSSSSPTSS-SSSSPSSPSSSLAESPEAA 157
 QY 65 KSETTIT---OSGYNGADVGGADNSTLETQNG--FRNNATTIQWNAKNSDI 113
 DB 158 GVSSTAPLPGGAAGPOTGVPAVSGALRELLACRNGDVSRRKRLVDAANVAKDM 212
 RESULT 10
 CSGB_ECOLI STANDARD; PRT; 151 AA.
 ID CSGB_ECOLI
 AC P39828;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Minor curli subunit precursor.
 GN CSGB OR B1041 OR Z1675 OR ECS1419.
 OS Escherichia coli, and
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 NX NCBI_TaxID=562, 83334;
 RX SEQUENCE FROM N.A.
 RP STRAIN=K12 / MCA100;
 RC MEDLINE=96414468; PubMed=8817489;
 RA Hammar M., Arngvist A., Bian Z., Olsen A., Normark S.;
 RT "Expression of two csg operons is required for production of
 RT fibronectin- and Congo red-binding curli polymers in Escherichia coli
 RT K-12.";
 RL Mol. Microbiol. 18:661-670(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97061202; PubMed=8905232;
 RA Oshima T., Alida H., Baba T., Fujita K., Hayashi K., Honjo A.,
 RA Ikemoto K., Ikeda T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
 RA Kimura S., Kitegawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
 RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horuchi T.;
 RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
 RT

RT corresponding to the 12.7-28.0 min region on the linkage map."
 RL DNA Res. 3:137-155(1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobeck E.J., Davis N.W., Lim A., Dimantanta E.T., Potamouis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
 RL Nature 409:529-533(2001).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,
 RA Kihara S., Shiba T., Hattori M., Shingawa H.;
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
 O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22(2001).
 RN [6]
 RP SEQUENCE OF 1-21 FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=95157246; PubMed=7851117;
 RA Argyris A., Olsen A., Normark S.;
 RT "Sigma S-dependent growth-phase induction of the csqA promoter in
 Escherichia coli can be achieved in vivo by sigma 70 in the absence
 of the nucleoid-associated protein H-NS.";
 RL Mol. Microbiol. 13:1021-1032(1994).
 CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
 CURLIN MONOMERS.
 CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL, X90754; CAA62281.1; -
 DR EMBL, AE000205; AAC74125.1; -
 DR EMBL, D90741; BA35831.1; -
 DR EMBL, AE005315; AAG55787.1; -
 DR EMBL, AP002554; BAB34842.1; -
 DR PIR, C90806; C90806.
 DR PIR, G85665; G85665.
 DR PIR, S70787; S70787.
 DR Ecogene: EGI2621; csGB.
 KW Fibrils; Signal; Complete proteome.
 FT SIGNAL
 FT CHAIN 1 21 POTENTIAL.
 FT SEQUENCE 151 AA; 15882 MW; B18D26B964014B8 CRC64;
 SQ
 Query Match 11.7%; Score 91.5; DB 1; Length 151;
 Best Local Similarity 29.1%; Pred. No. 0.48; Mismatches 54; Indels 7; Gaps 4;
 Matches 32; Conservative 17; Mismatches 54; Indels 7; Gaps 4;
 Oy 9 FAIVVSGSAL--AGVPPWGGGNNHNGGNSGSPDSTLSIY-QYSANALALQSDARK 65
 Db 33 FAVNELSKSFPNQAAII---GQAGTNNNSAQURGGSLTALVAVQESSNRA-KIDDTGY 88
 Oy 66 SETTIISGYNGADVQAGADNSTIELTONGFRNNATTIDQMAKNSDITV 115
 Db 89 NLAVYDQAGSANDASISQAGVNTAMITQKSGKANKITQYGTQKTAIV 138
 RESULT 11
 VQ38_BPT2 STANDARD; PRT; 262 AA.
 ID VQ38_BPT2
 AC P07875;
 DT 01-AUG-1998 (Rel. 08, Created)
 DT 01-AUG-1998 (Rel. 08, Last sequence update)
 DT 01-AUG-1990 (Rel. 15, Last annotation update)
 DE Receptor recognizing protein (Protein Gp38).
 GN 38.
 OS Bacteriophage T2.
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
 OC T4-like viruses.
 OC NCB1_TaxID=10664;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87283911; PubMed=3302276;
 RA Ride I., Drexler K., Eschbach M.L., Henning U.;
 RT "DNA sequence of genes 38 encoding a receptor-recognizing protein of
 bacteriophages T2, K3 and of K3 host range mutants.";
 RL J. Mol. Biol. 194:31-39(1987).
 CC -1- FUNCTION: Vg38 is at the tip of the long tail fibers and serves as
 the phage recognition site for the cellular receptor.
 CC -1- MISCELLANEOUS: THIS PHAGE USE OUTER MEMBRANE PROTEINS OMPF AND TTR
 AS RECEPTORS.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL, X05312; CAA28935.1; -
 DR PIR, S00275; S00275.
 DR InterPro: IPR007932; Tail_fibre_GP38.
 DR Pfam: PF05268; GP38; 1.
 KW Fiber protein; Phage recognition.
 SQ SEQUENCE 262 AA; 25801 MW; 0567366918F6C745 CRC64;
 Query Match 11.5%; Score 89.5; DB 1; Length 262;
 Best Local Similarity 34.1%; Pred. No. 1.3; Mismatches 33; Indels 17; Gaps 4;
 Matches 30; Conservative 8; Mismatches 33; Indels 17; Gaps 4;
 Oy 27 GGGGHHNGGNSGSPDSTLSIYQYSANALALQSDARKSETTIISGYNGADVQAGAD 86
 Db 175 GGGGHPFVGVGKIGSDSLTSGSVASL---TDAGTGTTF-QYGAGNGVNGAGCG 225
 Oy 87 NSTIELTONGFRNNATTIDQMAKNSDIT 114
 Db 226 -----RGWKKVYITSEGAAGAAYV 245
 RESULT 12
 MSA2_PLAF2 STANDARD; PRT; 347 AA.
 ID MSA2_PLAF2
 AC Q03646;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Merozoite surface antigen 2 precursor (MSA-2).
 GN MSA2.
 OS Plasmodium falciparum (isolate Ni932 / Nigeria).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OC NCB1_TaxID=70150;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91156685; PubMed=2000383;
 RA Smythe J.A., Coppel R.L., Day K.P., Martin R.K., Oduola A.M.T.,
 RA Kemp D.J., Anders R.F.;


```

ID OMB2_NEIMB STANDARD; PRT; 331 AA.
AC P30688;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Major outer membrane protein P.IB precursor (Protein IB) (PIB)
DE (Porin) (Class 3 protein).
GN FORB.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CCUG 37604 / M981 / Serogroup B / Serotype 4;
RX MEDLINE=93051225; PubMed=1330818;
RA Ward M.J., Lambden P.R., Heckels J.E.;
RT "Sequence analysis and relationships between meningococcal class 3
RT serotype proteins and other porins from pathogenic and non-pathogenic
RT Neisseria species."
RL FEMS Microbiol. Lett. 73:283-289 (1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CU385 / Serogroup B / Serotype 4 / Subtype 15;
RX MEDLINE=93116567; PubMed=1335540;
RA Zapata G.A., Vann W.F., Rubinstein Y., Frasch C.E.;
RT "Identification of variable region differences in Neisseria
RT meningitidis class 3 protein sequences among five group B
RT serotypes."
RL Mol. Microbiol. 6:3493-3499 (1992).
RN [3]
RP FUNCTION: Serves as a slightly cation selective porin.
CC -1- SUBUNIT: Homotrimer.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
CC -1- SIMILARITY: Belongs to the Gram-negative porin family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL, X65531; CAA46501.1; -
DR PIR, S21409; S21409.
DR PIR, S28441; S28441.
DR InterPro: IPR001702; Porin Gram-ve.
DR Pfam: PF00267; Gram-ve porins; 1.
DR PRINTS: PS00182; ECOLNEIPORIN.
DR PROSITE: PS00576; GRAM_NEG_PORIN, 1.
KW Outer membrane; Porin; Transmembrane; Antigen; Signal.
FT SIGNAL 1 19
FT CHAIN 20 331 MAJOR OUTER MEMBRANE PROTEIN P.IB.
FT VARIANT 70 71 NG -> KR (IN STRAIN CU385).
SQ SEQUENCE 331 AA; 35741 MW; 35FA35B7EBD28301 CRC64;
Query Match 11.2%; Score 87; DB 1; Length 331;
Best Local Similarity 25.5%; Pred. No. 2.8;
Matches 42; Conservative 23; Mismatches 78; Indels 22; Gaps 6;

```

```

RESULT 15
PER_DROWI STANDARD; PRT; 1093 AA.
ID PER_DROWI
AC Q03297; O18421; O18422; P91721; P91722;
DT 01-OCT-1993 (Rel. 27, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Period circadian protein (Fragment).
GN PER.
OS Drosophila willistoni (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7260;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Various strains;
RX MEDLINE=97357421; PubMed=9214747;
RA Gleason J.M., Powell J.R.;
RT "Interspecific and intraspecific comparisons of the period locus in
RT the Drosophila willistoni sibling species."
RL Mol. Biol. Evol. 14:741-753 (1997).
RN [2]
RP SEQUENCE OF 579-646 FROM N.A.
RX MEDLINE=93196482; PubMed=8450754;
RA Pelicoto A.A., Camposan S., Costa R.H., Kyriacou C.P.;
RT "Molecular evolution of a repetitive region within the per gene of
RT Drosophila."
RL Mol. Biol. Evol. 10:127-139 (1993).
CC -1- FUNCTION: Essential for biological clock functions. Determines the
CC period length of circadian and ultradian rhythms; an increase in
CC PER dosage leads to shortened circadian rhythms and a decrease
CC leads to lengthened circadian rhythms. Essential for the circadian
CC rhythmicity of locomotor activity, eclosion behavior, and for the
CC rhythmic component of the male courtship song that originates in
CC the thoracic nervous system. The biological cycle depends on the
CC complex light induces the degradation of TIM, which promotes
CC elimination of PER. Nuclear activity of the heterodimer
CC coordinatively regulates PER and TIM transcription through a
CC negative feedback loop. Behaves as a negative element in circadian
CC transcriptional loop. Does not appear to bind DNA, suggesting
CC indirect transcriptional inhibition (By similarity).
CC -1- SUBUNIT: Forms heterodimer with timeless (TIM); the complex then
CC translocates into the nucleus (By similarity).
CC -1- SUBCELLULAR LOCATION: Nuclear at specific periods of the day.
CC First accumulates in the perinuclear region about one hour before
CC translocation into the nucleus. Interaction with Tim is required
CC for nuclear localization (By similarity).
CC -1- PTM: PHOSPHORYLATED WITH A CIRCADIAN RHYTHMICITY, PROBABLY BY THE
CC DOUBLE-TIME PROTEIN (DBT). PHOSPHORYLATION COULD BE IMPLICATED IN
CC THE STABILITY OF PER MONOMER AND IN THE FORMATION OF HETERODIMER
CC PER-TIM (By similarity).
CC -1- MISCELLANEOUS: THE SEQUENCE SHOWN IS THAT OF STRAIN ATLIXCO.
CC -1- SIMILARITY: Contains 2 PAS (PER-ARNT-SIM) dimerization domain.
CC -1- SIMILARITY: Contains 1 PAS-associated C-terminal (PAC) domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL, U51055; AAB41360.1; -
DR EMBL, U51056; AAB41361.1; -
DR EMBL, U51057; AAB41362.1; -
DR EMBL, U51058; AAB41363.1; -
DR EMBL, U51059; AAB41364.1; -
DR EMBL, U51060; AAB41365.1; -
DR EMBL, U51061; AAB41366.1; -
DR EMBL, U51062; AAB41367.1; -

```

```

DR EMBL; U51063; AAB41368.1; -
DR EMBL; U51064; AAB41369.1; -
DR EMBL; U51065; AAB41370.1; -
DR EMBL; U51066; AAB41371.1; -
DR EMBL; U51067; AAB41372.1; -
DR EMBL; U51068; AAB41373.1; -
DR EMBL; U51069; AAB41374.1; -
DR EMBL; U51070; AAB41375.1; -
DR EMBL; U51071; AAB41376.1; -
DR EMBL; U51072; AAB41377.1; -
DR EMBL; L06342; AAA28765.1; -
DR FlyBase; FBgn0033161; Dwi1\per.
DR InterPro; IPR000014; PAS_domain.
DR Pfam; PF00989; PAS; 2.
DR SMART; SM00091; PAS; 2.
DR PROSITE; PS50112; PAS; 2.
KW Biological rhythms; Repeat; Nuclear protein; Phosphorylation;
KM Polymorphism.
FT NON_TER 1 1
FT DOMAIN <1 12 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 139 209 PAS 1.
FT DOMAIN 289 359 PAS 2.
FT DOMAIN 371 411 PAC.
FT DOMAIN 7 12 POLY-LYS.
FT DOMAIN 618 625 POLY-GLY.
FT DOMAIN 718 734 POLY-GLY.
FT DOMAIN 745 748 POLY-SER.
FT DOMAIN 759 770 POLY-GLY.
FT DOMAIN 885 888 POLY-ALA.
FT DOMAIN 911 917 POLY-ALA.
FT VARIANT T -> A (IN STRAIN 0811.4).
FT VARIANT 611 611 S -> F (IN STRAIN 0811.4).
FT VARIANT 617 617 G -> V (IN STRAIN 0811.4).
FT VARIANT 622 622 G -> A (IN STRAIN GUANA).
FT VARIANT 724 724 G -> S (IN STRAIN MANAUS 2).
FT VARIANT 726 726 MISSING (IN STRAIN SANTA MARIA).
FT VARIANT 729 734 MISSING (IN STRAIN PORTO ALEGRE 3).
FT VARIANT 730 734 MISSING (IN STRAIN PORTO ALEGRE 4).
FT VARIANT 731 734 MISSING (IN STRAINS MANAUS 1 AND MANAUS 3).
FT VARIANT 732 734 MISSING (IN STRAINS LIMA B, L/HABITATUE AND CANO MORA).
FT VARIANT 733 733 G -> V (IN STRAIN PORTO ALEGRE 4).
FT VARIANT 733 734 MISSING (IN STRAINS GUADELOUPE, MANAUS 2, PORTO ALEGRE 2, PORTO ALEGRE 1 AND GUANA).
FT VARIANT 734 734 MISSING (IN STRAINS MANAUS 4, PORTO ALEGRE 1 AND PORTO ALEGRE 2).
FT VARIANT 747 747 S -> A (IN STRAINS GUADELOUPE AND GUANA).
FT VARIANT 764 766 MISSING (IN STRAIN MANAUS 3).
FT VARIANT 886 886 A -> T (IN STRAIN 0811.4).
FT NON_TER 1093 1093
SQ SEQUENCE 1093 AA; 115896 MW; ABEDE050267EC187 CRC64;

Query Match 11.0%; Score 86; DB 1; Length 1093;
Best Local Similarity 27.2%; Pred. No. 13;
Matches 25; Conservative 6; Mismatches 45; Indels 16; Gaps 2;

QY 27 GGGGNNHGGGNSGSPDSTLSIYOGSANAALALQSDARKSETTITSGYNGADYVQCAD 86
DB 721 GGGGGGGGGGGGGGLPLFLVTHITS-----SSQNKGPPTGVAGAGAGGGGGGG-- 770
QY 87 NSTIELTONGFRNNATIDQWNAKNSDIYVQY 118
DB 771 -----SCSGLGNGNNGVSGNGNNSQSPSTNOY 796

```

Search completed: March 11, 2004, 18:34:52
Job time : 7.3 secs

THIS PAGE BLANK (USP)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Comugen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:23:54 ; Search time 30.5 Seconds
(without alignments)
1562.074 Million cell updates/sec

Title: US-09-543-407-12

Perfect score: 779
Sequence: 1 MKLLKVAAPFAIVSGSALA.....VTRVTHEMAHANNATANOY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapex 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_prodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriaph:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	686	88.1	152	2	033802
2	591.5	75.9	150	2	07X243
3	537	68.9	149	2	07X240
4	495.5	63.6	152	16	08CW63
5	427.5	49.4	150	2	07X237
6	385	45.9	76	2	054069
7	122	14.6	29	2	09S305
8	113.5	14.6	3501	16	08Y106
9	113.5	14.6	3552	16	08XSD6
10	108.5	13.9	191	10	07XDR3
11	107	13.7	502	16	08EIR4
12	106	13.6	1748	5	094821
13	105	13.5	2035	2	09XCJ4
14	105	13.5	2039	16	08ZNS7
15	104.5	13.4	1286	2	0841Y5
16	104	13.4	2174	16	092U08

17	103	13.2	139	16	08EIH3	08eih3 shewanella
18	102.5	13.2	624	3	08N1V1	08n1v1 neurospora
19	102	13.1	1422	16	08EFU3	08efu3 shewanella
20	100.5	12.9	152	2	07X241	07x241 citrobacter
21	100	12.8	179	2	033801	033801 salmonella
22	99.5	12.8	151	2	07X244	07x244 citrobacter
23	99	12.7	1765	16	07V8S5	07v8s5 prochloroco
24	98.5	12.6	1222	5	09M4F0	09m4f0 drosophila
25	98.5	12.6	1321	5	08T8L9	08t8l9 drosophila
26	98.5	12.6	1351	5	08SX56	08sx56 drosophila
27	98.5	12.6	1354	5	08MPN4	08mpn4 drosophila
28	98.5	12.6	7716	16	07WZ28	07wz28 rhodospirilla
29	97.5	12.5	154	16	08SJT5	08sjt5 bradyrhizob
30	97.5	12.5	348	13	09J397	09j397 cyprinus ca
31	97.5	12.5	739	2	09X687	09x687 salmonella
32	97.5	12.5	3659	16	098LN6	098ln6 rhizobium 1
33	97	12.5	1618	2	09KRB1	09krb1 rickettsia
34	96.5	12.4	151	2	07X238	07x238 enterobacte
35	96	12.3	145	16	08U6N9	08u6n9 agrobacteri
36	96	12.3	157	16	08BHG0	08bhg0 pseudomonas
37	95.5	12.3	130	16	089J14	089j14 bradyrhizob
38	95.5	12.3	453	5	09N6M8	09n6m8 drosophila
39	95.5	12.3	1615	2	09KXA8	09kxa8 rickettsia
40	95	12.2	362	16	08EV84	08ev84 mycoplasma
41	95	12.2	163	2	084U05	084u05 rickettsia
42	95	12.2	1618	2	09KXB4	09kxb4 escherichia
43	94.5	12.1	160	16	08CM64	08cm64 escherichia
44	94.5	12.1	453	5	09NGF6	09ngf6 drosophila
45	94.5	12.1	453	5	09NGF7	09ngf7 drosophila

ALIGNMENTS

RESULT 1

033802 ID 033802 PRELIMINARY; PRT; 152 AA.
AC 033802;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Agfa protein (Fragment).
GN AGFA.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98053981; PubMed=9393832;
RA Sukupolvi S.S., Lorentz R.G., Gordon J.I., Bian Z., Pfeiffer J.D.,
RA Normark S.J., Rhen M.;
RT "Expression of thin, aggregative fimbriae promotes interaction of
RT Salmonella typhimurium SR-11 with mouse small intestinal epithelial
RT cells.";
RL Infect. Immun. 65:5320-5325 (1997).
DR EMBL; AJ000514; CAA04151.1; -;
FT NON_TER 152
SQ SEQUENCE 152 AA; 15401 MW; 9DA7DADC2364B006 CRC64;

Query Match 88.1%; Score 686; DB 2; Length 152;
Best local Similarity 88.7%; Pred. No. 4.1e-45;
Matches 134; Conservative 5; Mismatches 12; Indels 0; Gaps 0;

Qy	1	MKLLKVAAPFAIVSGSALA	GVPPQGGGNNHNGGNSG	SPDSTLSTIYQYGS	NNAAALAQ 60
Db	1	MKLLKVAAPFAIVSSAAV	GVPPQGGGNNHNGGNSG	SPDSTLSTIYQYGS	NNAAALAQ 60
Qy	61	SDARKSETTITOSGYNGA	DVGADN	STIELTQNGFRNNATID	QNNAKSDITTVQYGG 120
Db	61	SDARKSETTITOSGYNGA	DVGADN	STIELTQNGFRNNATID	QNNAKSDITTVQYGG 120
Qy	121	NNAAALVNDQLVTRVVT	THEMAHANNATANOY 151		

Db 121 NNAALVNDQVTRVVTHEMAHANNATANOY 151

RESULT 2

QY 07X243 PRELIMINARY; PRT; 150 AA.
AC 07X243;
DT 01-OCT-2003 (TREMblrel. 25, Created)
DT 01-OCT-2003 (TREMblrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)
DE Curlin-csgA protein.
GN CSGA.
OS Citrobacter sp. Fec2.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Citrobacter.
OX NCBI_TaxID=213763;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fec2;
RA Zogaj X., Bokranz W., Nimtz M., Romling U.;
RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RT Enterobacteriaceae Isolated from the Human Gastrointestinal Tract."
RL Infect. Immun. 72:4151-4158 (2003).
DR EMBL: AJ515700; CADS6672.1; -
SQ SEQUENCE 150 AA; 15016 MW; 1D7141B8D6973DC6 CRC64;

Query Match 75.9%; Score 591.5; DB 2; Length 150;
Best Local Similarity 77.5%; Pred. No. 6,7e-38;
Matches 117; Conservative 14; Mismatches 19; Indels 1; Gaps 1;

QY 1 MKLKVAAFAIIVSGSALAGVPPQWGGGNNHNGSGSPDSTLSIYOGSANAALALQ 60
Db 1 MKLKVAAFAIIVSGSALAGVPPQWGGGNNHNGSGSPDSTLSIYOGSANAALALQ 59
QY 61 SDARKSETTITGSGYNGADVGQADNSTIETLTONGFRNNATIDOWNAKNSDITVQYGG 120
Db 60 SDARKSDTTITGQRFNGADVGQGSNDSTIDLTQRFNKNATIDOWNKNSDITVQYGG 119
QY 121 NNAALVNDQVTRVVTHEMAHANNATANOY 151
Db 120 NNAALVNDQVTRVVTHEMAHANNATANOY 150

RESULT 3

QY 07X240 PRELIMINARY; PRT; 149 AA.
AC 07X240;
DT 01-OCT-2003 (TREMblrel. 25, Created)
DT 01-OCT-2003 (TREMblrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)
DE Curlin-csgA protein.
GN CSGA.
OS Citrobacter freundii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Citrobacter.
OX NCBI_TaxID=546;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fec4;
RA Zogaj X., Bokranz W., Nimtz M., Romling U.;
RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RT Enterobacteriaceae Isolated from the Human Gastrointestinal Tract."
RL Infect. Immun. 72:4151-4158 (2003).
DR EMBL: AJ515701; CADS6675.1; -
SQ SEQUENCE 149 AA; 15260 MW; 946D52017F648FD CRC64;

Query Match 68.9%; Score 537; DB 2; Length 149;
Best Local Similarity 71.5%; Pred. No. 9,7e-34;
Matches 108; Conservative 17; Mismatches 24; Indels 2; Gaps 1;

QY 1 MKLKVAAFAIIVSGSALAGVPPQWGGGNNHNGSGSPDSTLSIYOGSANAALALQ 60
Db 1 MKLKVAAFAIIVSGSALAGVPPQWGGGNNHNGSGSPDSTLSIYOGSANAALALQ 60

Db 1 MKLKVAAFAIIVSGSALAGVPPQW--GGNHGGSNYGPDSTLSIYOGSNNANALQ 58

QY 61 SDARKSETTITGSGYNGADVGQADNSTIETLTONGFRNNATIDOWNAKNSDITVQYGG 120
Db 59 SDARKSDTTITGQRFNGADVGQGSNDSTIDLTQRFNKNATIDOWNKNSDITVQYGG 118

QY 121 NNAALVNDQVTRVVTHEMAHANNATANOY 151
Db 119 NNAALVNDQVTRVVTHEMAHANNATANOY 149

RESULT 4

QY 08CW63 PRELIMINARY; PRT; 152 AA.
AC 08CW63;
DT 01-MAR-2003 (TREMblrel. 23, Created)
DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
DE Major curlin subunit precursor.
GN CSGA OR C1306.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O6:HI / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G., III, Redford P., Roesech P.,
RA Raske D., Buckles E.L., Liou S.-R., Boutin A., Hackert J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Petina N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blatner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli."
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).
DR EMBL: AE016759; AAN79779.1; -
KW Complete proteome.
SQ SEQUENCE 152 AA; 15064 MW; 3BA57F34C1240E83 CRC64;

Query Match 63.6%; Score 495.5; DB 16; Length 152;
Best Local Similarity 65.1%; Pred. No. 1.5e-30;
Matches 99; Conservative 21; Mismatches 31; Indels 1; Gaps 1;

QY 1 MKLKVAAFAIIVSGSALAGVPPQW--GGNHGGSNYGPDSTLSIYOGSANAALALQ 59
Db 1 MKLKVAAFAIIVSGSALAGVPPQWGGGNNHNGSGSPDSTLSIYOGSANAALALQ 60
QY 60 SDARKSETTITGSGYNGADVGQADNSTIETLTONGFRNNATIDOWNAKNSDITVQYGG 119
Db 61 QADARNSDLTITGQRFNGADVGQGSNDSTIDLTQRFNKNATIDOWNKNSDITVQYGG 120
QY 120 NNAALVNDQVTRVVTHEMAHANNATANOY 151
Db 121 GNGAALVNDQVTRVVTHEMAHANNATANOY 152

RESULT 5

QY 07X237 PRELIMINARY; PRT; 150 AA.
AC 07X237;
DT 01-OCT-2003 (TREMblrel. 25, Created)
DT 01-OCT-2003 (TREMblrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)
DE Curlin-csgA protein.
GN CSGA.
OS Enterobacter sakazakii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Enterobacter.
OX NCBI_TaxID=28141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fec39;
RA Zogaj X., Bokranz W., Nimtz M., Romling U.;

QY 122 N 122
DB 350 N 350

RESULT 12

Q94821 PRELIMINARY; PRT; 1748 AA.
ID Q94821
AC Q94821; P92145; P92144; P92143; P92142; P92141; Q94820;
DT 01-FEB-1997 (TEMBLrel. 02, Created)
DT 01-FEB-1997 (TEMBLrel. 02, Last sequence update)
DE 01-OCT-2003 (TEMBLrel. 25, Last annotation update)
DN CNUB protein.
OS Tetrahymena thermophila.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
OC Tetrahymena; Tetrahymena.
OX NCBI_TaxID=5911;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88189811; PubMed=357771;
RA Martindale D.W., Taylor F.M.;
RT "Multiple introns in a conjugation-specific gene from Tetrahymena thermophila.";
RL Nucleic Acids Res. 16:2189-2201(1988).
DR EMBL: L03710; AAC3171.1; -;
DR PIR: S42136; S42136.
DR HSP; P05888; IAAF.
DR GO: GO:0003676; F:nucleic acid binding; IEA.
DR InterPro: IPR001878; Znf_CCHC.
DR Pfam: PF00098; Zf_CCHC; 7.
DR PRINTS: PR00939; C2HCNFINGER.
DR SMART: SM00343; Znf_CCHC; 7.
DR PROSITE: PS50158; Zf_CCHC; 7.
FT CONFLICT 251 251 M -> I (IN REF. 1).
FT CONFLICT 256 256 I -> N (IN REF. 1).
SQ SEQUENCE 1748 AA; 199624 MW; 0B03F210104008A3 CRC64;
Query Match 13.6%; Score 106; DB 5; Length 1748;
Best Local Similarity 32.1%; Pred. No. 14;
Matches 35; Conservative 13; Mismatches 33; Indels 28; Gaps 5;
DB 25 QMGCGGNHNGG---GNSSGPDSTLSTIYGSANALALQSDARKSETTIT---OSGYGN 77
ID Q94821 PRELIMINARY; PRT; 2035 AA.
AC Q94821;
DT 01-NOV-1999 (TEMBLrel. 12, Created)
DT 01-NOV-1999 (TEMBLrel. 12, Last sequence update)
DE 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
DN SHDA.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RESULT 13

RP SEQUENCE FROM N.A.
RC STRAIN=ATCC14028;
RA Kingley R.A., van Amsterdam K., Baumber A.J.;
RT "The presence of a pathogenicity island specific to Salmonella enterica subspecies I correlates with adaptation to warm blooded animals.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC14028;
RA Kingley R.A., van Amsterdam K., Edwards E.W., Hargis B.M., Baumber A.J.;
RT "Complete sequence of the xseA-his intergenic region of the S. enterica serotype Typhimurium genome and its distribution within the genus Salmonella.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF140550; AAD25110.2; -;
DR InterPro: IPR006315; Autotransport.
DR InterPro: IPR005546; Autotransporter.
DR InterPro: IPR004899; Pectactin.
DR Pfam: PF03797; Autotransporter; 1.
DR Pfam: PF03212; Pectactin; 1.
DR TIGRfam: TIGR01414; autoclans_bar1; 3.
DR PROSITE: PS00584; PFK_KINASES_2; 2.
SQ SEQUENCE 2035 AA; 207032 MW; 295DB82FEC84FAB CRC64;

Query Match 13.5%; Score 105; DB 2; Length 2035;
Best Local Similarity 26.3%; Pred. No. 20;
Matches 54; Conservative 20; Mismatches 61; Indels 70; Gaps 11;

QY 10 AAIWVGSAALAG-----VWPQWCGGNGHNGG-----NSSGPD----- 42
DB 90 AALYVSGVATVGMQPTVTGTGIVETSGGADDPGKRVSAISLDHAIILETDAKI 149
QY 43 STLSIYQGSANA-----ALAIQSPARKS-ETITQSGYGNADYG 82
DB 150 TTGTGYTGISAAADGTLRLTDTLTIDGNFGVMTLYTSGEATLDTIVEANSSAOYO 209
QY 83 QGA-----DNSTIELQ---NGFRNATIDOWNAXY-SQITVQVY-----GNNALV 127
DB 210 QGSSTLNVLDGSTITTLAQGINVAGTATDGGSTLNLSDSSVSSAGTMTIGCTNOALN 269
QY 128 YDQLVTRVVTHEVA-----HANNAT 147
DB 270 ---LTNATITHTNAGAAVANNAT 291

RESULT 14
Q82N57 PRELIMINARY; PRT; 2039 AA.
ID Q82N57;
AC Q82N57;
DT 01-MAR-2002 (TEMBLrel. 20, Created)
DT 01-MAR-2002 (TEMBLrel. 20, Last sequence update)
DE 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
DE Similar to the C-terminal region of AIDA, IcsA, subspecies I specific, Peyer's patch colonization and shedding factor.
DN SHDA OR STM2513.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.B., Spieth J., Clifton S.W., Latreille P., Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D., Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E., Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M., Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium LT2.";

RL Nature 413:852-856(2001).
 DR EMBL: AE008813; AAL21407.1; -.
 DR InterPro: IPR006315; Autotransport.
 DR InterPro: IPR005546; Autotransporter.
 DR InterPro: IPR004899; Peractin.
 DR InterPro: IPR002173; PfkB.
 DR Pfam: PF03797; Autotransporter; 1.
 DR Pfam: PF03212; Peractin; 1.
 DR TIGRfam: TIGR01414; autotrans part; 3.
 DR PROSITE: PS00584; PFKB_KINASES_2; 2.
 KW Complete proteome.
 SQ SEQUENCE 2039 AA; 207127 MW; 894E1F8F29339BA CRC64;
 Query Match 13.5%; Score 105; DB 16; Length 2039;
 Best Local Similarity 26.3%; Pred. No. 20;
 Matches 54; Conservative 20; Mismatches 61; Indels 70; Gaps 11;
 QY 10 AAIIVSGSALAG-----VVPQWGGGNNHNGG-----NSSGPD----- 42
 DB 94 AALVYSGVATVGMQPTVTGTGVIETSGGADPDGKGVSNALSLDHYAILLELTDKI 153
 QY 43 STLSIYOGSANA-----AALQSDARKS-ETTTGSGYGNADYV 82
 DB 154 TTGTGTYGCTISADGSLRLDSTLTIDGNGFVMTLYGSEXTLDTGIVEANSSAQVQ 213
 QY 83 QGA-----DNSTIELTQ--NGFRNNATIDQWNAKN-SDITVGYG-----GNVAALVN 127
 DB 214 QGSTLNVLDSGTTTLAQGQINVAAGNTATDEGSTLNLSDSVSSAGMTSTIGTKKALN 273
 QY 128 YQQLTRVVTHEMA-----HANNAT 147
 DB 274 ---LTNATITHTNAGAAVQANNAT 295
 RESULT 15
 Q841Y5 PRELIMINARY; PRT; 1286 AA.
 ID Q841Y5;
 AC 01-UN-2003 (TREMBlrel. 24, Created)
 DT 01-UN-2003 (TREMBlrel. 24, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Putative high-molecular-weight surface-exposed protein Cf0009.
 OS Campylobacter fetus.
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
 OC Campylobacteriaceae; Campylobacter.
 OX NCBI_TaxID=196;
 RX MEDLINE=90354448; PubMed=2387868;
 RA Blaser M.J., Gotschlich E.C.;
 RT "Surface array protein of Campylobacter fetus. Cloning and gene structure."
 RL J. Biol. Chem. 265:14529-14535(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=23D;
 RX MEDLINE=91035477; PubMed=2229082;
 RA Blaser M.J., Gotschlich E.C.;
 RT "Surface array protein of Campylobacter fetus. Cloning and gene structure."
 RL J. Biol. Chem. 265:19372-19372(1990).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=23D;
 RX MEDLINE=92394895; PubMed=1522068;
 RA Tummuru M.K., Blaser M.J.;
 RT "Characterization of the Campylobacter fetus sapA promoter: evidence that the sapA promoter is deleted in spontaneous mutant strains."
 RL J. Bacteriol. 174:5916-5922(1992).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=23D;

RX MEDLINE=93348254; PubMed=8346244;
 RA Tummuru M.K., Blaser M.J.;
 RT "Rearrangement of sapA homologs with conserved and variable regions in Campylobacter fetus."
 RL Proc. Natl. Acad. Sci. U.S.A. 90:7265-7269(1993).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=23D;
 RX MEDLINE=95204338; PubMed=7896695;
 RA Dworkin J., Tummuru M.K., Blaser M.J.;
 RT "A lipopolysaccharide-binding domain of the Campylobacter fetus S-layer protein resides within the conserved N terminus of a family of silent and divergent homologs."
 RL J. Bacteriol. 177:1734-1741(1995).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN=23D;
 RX MEDLINE=99069317; PubMed=9851986;
 RA Thompson S.A., Shedd O.L., Ray K.C., Beins M.H., Jorgensen J.P., Blaser M.J.;
 RT "Campylobacter fetus surface layer proteins are transported by a type I secretion system."
 RL J. Bacteriol. 180:6450-6458(1998).
 RN [7]
 RP SEQUENCE FROM N.A.
 RC STRAIN=23D;
 RX PubMed=12694614;
 RA Tu Z.C., Wassenaar T.M., Thompson S.A., Blaser M.J.;
 RT "Structure and genotypic plasticity of the Campylobacter fetus sap locus."
 RL Mol. Microbiol. 48:685-698(2003).
 DR EMBL: AY211269; AAO64216.1; -.
 DR InterPro: IPR005546; Autotransporter.
 DR Pfam: PF03797; Autotransporter; 1.
 SQ SEQUENCE 1286 AA; 134079 MW; A1F9C8C34158789 CRC64;
 Query Match 13.4%; Score 104.5; DB 2; Length 1286;
 Best Local Similarity 27.8%; Pred. No. 13;
 Matches 44; Conservative 26; Mismatches 57; Indels 31; Gaps 9;
 QY 5 KYAAFAIIVSGSALAGVVPQWGGGNNHNGGNS---SGPDSTLSIYOGSANAALAQ 61
 DB 376 QVASENLVLIISGTTN--VPTIGGSGATVATNNQVLTISGKVTSTIYGNNAN-----K 427
 QY 62 DAKSETTTIGSGYGNADV--GQGANSTIELTQNGFRNNATIDQWNAKNSDITVGYG 119
 DB 428 SANENKVTITF-GTANVADIDYGGKSIINNSI-----ANKSITISGGTLQVTNI--YG 477
 QY 120 GNNAAALVNYDQ-----VTRVV-THENAHANNATAN 149
 DB 478 GHSAAKQANENSIGISNGNINNIYVGHQAQDHTNLTNTIN 515

Search completed: March 11, 2004, 18:40:11
 Job time : 32.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:13:53 ; Search time 45.9 Seconds
(without alignments)
929,514 Million cell updates/sec

Title: US-09-543-407-14

Perfect score: 775
Sequence: 1 MGLKLVAAFAIVSGSALA.....HASVWRYQVFGNNATATNOY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq_29Jan04:*

- 1: genebegp19808:*
- 2: genebegp19908:*
- 3: genebegp20008:*
- 4: genebegp20018:*
- 5: genebegp20028:*
- 6: genebegp20038:*
- 7: genebegp20038s:*
- 8: genebegp20048:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	775	100.0	151	3	AAB36347
2	714	92.1	151	3	AAB36352
3	696	89.8	151	3	AAB36346
4	693	89.4	151	2	AAB74625
5	683	89.4	151	3	AAB36341
6	682	89.3	151	2	AAB23570
7	659	85.0	151	3	AAB36355
8	612	79.0	151	3	AAB36353
9	611	78.8	151	3	AAB36349
10	609	78.6	151	3	AAB36350
11	603	77.8	151	3	AAB36354
12	602	77.7	151	3	AAB36351
13	578	74.6	151	3	AAB36348
14	528	68.1	151	3	AAB36343
15	523	67.5	151	7	ABR82651
16	514	66.3	120	2	AAB62761
17	514	66.3	120	2	AAB23569
18	463	59.7	142	2	AAB52664
19	391	50.5	122	2	AAB52663
20	237	30.6	45	3	AAB36316
21	132	17.0	22	3	AAB36318
22	123	15.9	23	3	AAB36321
23	123	15.9	23	3	AAB36326
24	123	15.9	23	3	AAB36338
25	115	14.8	22	3	AAB36325

26	115	14.8	22	3	AAB36339	Aab36339	Salmonell
27	115	14.8	22	3	AAB36320	Aab36320	Salmonell
28	113	14.6	24	7	ABR82644	Abbr82644	E. coli c
29	109	14.1	23	3	AAB36340	Aab36340	Salmonell
30	109	14.1	23	3	AAB36324	Aab36324	Salmonell
31	109	14.1	23	3	AAB36319	Aab36319	Salmonell
32	104.5	13.5	151	3	AAB36344	Aab36344	Escherich
33	98.5	12.7	151	3	AAB36342	Aab36342	Salmonell
34	98	12.6	26	7	ABR82645	Abbr82645	E. coli c
35	95	12.3	186	6	ABU21488	Abu21488	Protein e
36	93	12.0	19	3	AAB36323	Aab36323	Salmonell
37	93	12.0	19	3	AAB36336	Aab36336	Salmonell
38	93	12.0	19	3	AAB36328	Aab36328	Salmonell
39	92	11.9	23	3	AAB36331	Aab36331	Escherich
40	92	11.9	673	5	AAy44403	Aay44403	Human tru
41	92	11.9	673	5	AAU79538	Aau79538	Truncated
42	92	11.9	949	5	AAy44404	Aay44404	Human tru
43	92	11.9	949	5	AAU79539	Aau79539	Truncated
44	92	11.9	1327	3	AAy44402	Aay44402	Human tan
45	92	11.9	1327	3	AAB27212	Aab27212	Human tan

ALIGNMENTS

RESULT 1		
AAB36347		
ID	AAB36347	standard; protein; 151 AA.
XX	XX	XX
AC	AAB36347;	
XX	XX	XX
DT	26-FEB-2001	(first entry)
XX	XX	XX
DE	Agfa::PT3#2 amino acid sequence	SEQ ID NO:14.
XX	XX	XX
KW	Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;	
KW	vacine; immune response; immunogen.	
XX	XX	XX
OS	Salmonella enteritidis.	
OS	Escherichia coli.	
XX	Synthetic.	
XX	XX	XX
PN	WO20060102-A2.	
XX	XX	XX
PD	12-OCT-2000.	
XX	XX	XX
PF	05-APR-2000; 2000WO-CA000356.	
XX	XX	XX
PR	05-APR-1999; 99US-0127888P.	
XX	XX	XX
PA	(UYVI-) UNIV VICTORIA.	
XX	XX	XX
PI	White AP, Doran JL, Collison SK, Kay MW;	
XX	XX	XX
DR	WPI; 2000-672631/65.	
XX	XX	XX
DR	N-PSDB; AAC64623.	
XX	XX	XX
PT	Recombinant agfa gene having a segment replaced by a foreign DNA sequence	
PT	which encodes foreign epitope or antigen, expresses recombinant Agfa	
XX	protein useful for eliciting immune response in animal.	
XX	XX	XX
PS	Disclosure; Page 136; 139pp; English.	
XX	XX	XX
CC	The present invention describes a recombinant agfa gene (1) where a	
CC	segment of the gene has been replaced by a segment of a foreign DNA	
CC	sequence which encodes a foreign epitope or antigen. Also described are:	
CC	(1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended	
CC	assembly system of strains of Salmonella, Escherichia coli and	
CC	Enterobacteriaceae for the production of fimbriae comprising recombinant	
CC	Agfa, GsgA and Agfa-homologue fimbria subunits, respectively; (2)	
CC	directing recombination of a recombinant gene into the chromosome of the	
CC	homologous species; (3) directing recombination of a recombinant gene	
CC	back into the chromosome of the homologous species, replacing the native	

CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent; (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

XX Sequence 151 AA;

Query Match 100.0%; Score 775; DB 3; Length 151;
 Best Local Similarity 100.0%; Pred. No. 5, 4e-67;
 Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLKVAAPFAIVSGSALAGVPPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
 DB 1 MKLLKVAAPFAIVSGSALAGVPPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
 QY 61 SDARKSETTITSGYGNAGDVGGAGNNTIELTONGFRNNATTIDOMNAKNSDITVGQYDQ 120
 DB 61 SDARKSETTITSGYGNAGDVGGAGNNTIELTONGFRNNATTIDOMNAKNSDITVGQYDQ 120
 QY 121 LVTRVVTHEMAHSAVMVRQVGFNNATANQY 151
 DB 121 LVTRVVTHEMAHSAVMVRQVGFNNATANQY 151

RESULT 2

AAB36352 standard; protein; 151 AA.

AC AAB36352;
 DT 26-FEB-2001 (first entry)

DE Agfa::PT3#7 amino acid sequence SEQ ID NO:24.

KW *Salmonella*; agfa; chromosomal gene replacement; fimbria; epitope;
 KM vaccine; immune response; immunogen.

OS *Salmonella enteritidis*.

OS *Escherichia coli*.

OS Synthetic.

PN WO200060102-A2.

PD 12-OCT-2000.

PF 05-APR-2000; 2000WO-CA000356.

PR 05-APR-1999; 99US-0127888P.

PA (UYVI-) UNIV VICTORIA.

PI White AP, Doran JL, Collison SK, Kay WW;
 DR WPI; 2000-672631/65.
 DR N-PSDB; AAC64628.

PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.

PS Disclosure; Page 138; 139pp; English.

XX The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SBF17/TAF) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

XX Sequence 151 AA;

Query Match 92.1%; Score 714; DB 3; Length 151;
 Best Local Similarity 91.1%; Pred. No. 4, 3e-61;
 Matches 144; Conservative 0; Mismatches 0; Indels 14; Gaps 2;

QY 1 MKLLKVAAPFAIVSGSALAGVPPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
 DB 1 MKLLKVAAPFAIVSGSALAGVPPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
 QY 61 SDARKSETTITSGYGNAGDVGGAGNNTIELTONGFRNNATTIDOMNAKNSDITVGQYDQ 120
 DB 61 SDARKSETTITSGYGNAGDVGGAGNNTIELTONGFRNNATTIDOMNAKNSDITVGQYDQ 120
 QY 121 LVTRVVTHEMAHSAVMVRQVGFNNATANQY 151
 DB 114 LVTRVVTHEMAHSAVMVRQVGFNNATANQY 151

RESULT 3

AAB36346 standard; protein; 151 AA.

AC AAB36346;
 DT 26-FEB-2001 (first entry)

DE Agfa::PT3#1 amino acid sequence SEQ ID NO:12.

KW *Salmonella*; agfa; chromosomal gene replacement; fimbria; epitope;
 KM vaccine; immune response; immunogen.

OS *Salmonella enteritidis*.

OS *Escherichia coli*.

OS Synthetic.

PN WO200060102-A2.

PD 12-OCT-2000.

PF 05-APR-2000; 2000WO-CA000356.

PR 05-APR-1999; 99US-0127888P.

PA (UYVI-) UNIV VICTORIA.

XX White AP, Doran JL, Collison SK, Kay WW;
 XX WPI; 2000-672631/65.
 DR N-PSDB; AAC64622.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 135; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SFF1/TAFF) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
 CC Agfa, CagA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal.
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;
 Query Match 89.8%; Score 696; DB 3; Length 151;
 Best Local Similarity 87.6%; Pred. No. 2,4e-59;
 Matches 141; Conservative 0; Mismatches 0; Indels 20; Gaps 2;
 QY 1 MKLLKTAAPAAIYVSSGALAGVVPQWGGGNNHGGNSGSPDSTLSIYQYGSANPAALALQ 60
 DB 1 MKLLKTAAPAAIYVSSGALAGVVPQWGGGNNHGGNSGSPDSTLSIYQYGSANPAALALQ 60
 QY 61 SDARKSETTITGSGYNGADVGGADNSTIELTQNGFRNNATTIDQNNAKNSDITVQYQ-- 117
 DB 61 SDARKSETTITGSGYNGADVGGADNSTIELTQNGFRNNATTIDQNNAKNSDITVQYQ 120
 QY 118 -----YDQLVRYVYTHEMAHASVMWRYGVFGNNATTANQY 151
 DB 121 NNAAALVYDQLVTRVYVTHEMAHA-----NNATTANQY 151
 RESULT 4
 AAR74625
 ID AAR74625 standard; protein; 151 AA.
 AC AAR74625;
 XX
 DT 25-MAR-2003 (revised)
 DT 26-JUN-1995 (first entry)
 XX
 DE Agfa sequence.
 XX
 KM *Salmonella*; Agfa; vaccine.
 XX
 OS *Salmonella*.
 XX

PN W09425598-A2.
 XX
 PD 10-NOV-1994.
 XX
 PF 26-APR-1994; 94WO-IB000207.
 XX
 PR 26-APR-1993; 93US-00054452.
 XX
 PA (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
 PA (KING/) KING J.
 XX
 PI Kay WW, Collinson SK, Clouthier SC, Doran JL;
 XX WPI; 1994-358275/44.
 DR N-PSDB; AAO87467.
 XX
 PT Eliciting an immune response to *Salmonella* - using attenuated *Salmonella*
 PT strains, vector constructs, or compens. contg. fimbrial type proteins.
 XX
 PS Disclosure; Fig 7B; 95pp; English.
 XX
 CC The *Salmonella* Agfa protein and DNA are used in vaccine and genetic
 CC immunization compositions, respectively, to elicit an immune response to
 CC *Salmonella* in animals (e.g. food producing animals) and humans. (Updated
 CC on 25-MAR-2003 to correct PN field.)
 CC
 XX
 SQ Sequence 151 AA;
 Query Match 89.4%; Score 693; DB 2; Length 151;
 Best Local Similarity 90.7%; Pred. No. 4,7e-59;
 Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;
 QY 1 MKLLKTAAPAAIYVSSGALAGVVPQWGGGNNHGGNSGSPDSTLSIYQYGSANPAALALQ 60
 DB 1 MKLLKTAAPAAIYVSSGALAGVVPQWGGGNNHGGNSGSPDSTLSIYQYGSANPAALALQ 60
 QY 61 SDARKSETTITGSGYNGADVGGADNSTIELTQNGFRNNATTIDQNNAKNSDITVQYQ 120
 DB 61 SDARKSETTITGSGYNGADVGGADNSTIELTQNGFRNNATTIDQNNAKNSDITVQYQ 120
 QY 121 LVTRVYVYTHEMAHASVMWRYGVFGNNATTANQY 151
 DB 121 NNAAALVYDQLVTRVYVTHEMAHA-----NNATTANQY 151
 RESULT 5
 AAB36341
 ID AAB36341 standard; protein; 151 AA.
 AC AAB36341;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE *Salmonella* enteritidis Agfa amino acid sequence SEQ ID NO:5.
 XX
 KM *Salmonella*; agfa; chromosomal gene replacement; fimbrial; epitope;
 KM vaccine; immune response; immunogen.
 XX
 OS *Salmonella* enteritidis.
 XX
 PN W0200060102-A2.
 PN W0200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PP 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 PA
 XX
 PI White AP, Doran JL, Collison SK, Kay WW;
 XX WPI; 2000-672631/65.
 DR

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:13:53 ; Search time 45.9 Seconds
(without alignments)
929.514 Million cell updates/sec

Title: US-09-543-407-14

Perfect score: 775
Sequence: 1 MLLKVAAPAAIVSGSALA.....HASVWVQVFGNNATANY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_293Jan04:*

1: geneseqp19608:*

2: geneseqp19908:*

3: geneseqp20008:*

4: geneseqp20018:*

5: geneseqp20028:*

6: geneseqp20038:*

7: geneseqp20038:*

8: geneseqp20048:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	775	100.0	151	3	AAB36347
2	714	92.1	151	3	AAB36352
3	696	89.8	151	3	AAB36346
4	693	89.4	151	2	AAR74625
5	693	89.4	151	3	AAB36341
6	692	89.3	151	2	AAW23570
7	659	85.0	151	3	AAB36355
8	612	79.0	151	3	AAB36353
9	611	78.8	151	3	AAB36349
10	609	78.6	151	3	AAB36350
11	603	77.7	151	3	AAB36354
12	602	77.7	151	3	AAB36351
13	578	74.6	151	3	AAB36348
14	528	68.1	151	3	AAB36343
15	523	67.5	151	7	ABR82651
16	514	66.3	120	2	AAW23569
17	514	66.3	120	2	AAW23569
18	463	59.7	142	2	AAR52664
19	391	50.5	122	2	AAR52663
20	237	30.6	45	3	AAB36316
21	132	17.0	22	3	AAB36318
22	123	15.9	23	3	AAB36321
23	123	15.9	23	3	AAB36326
24	123	15.9	23	3	AAB36328
25	115	14.8	22	3	AAB36325

26	115	14.8	22	3	AAB36339	Aab36339
27	115	14.8	22	3	AAB36320	Aab36320
28	113	14.6	24	7	ABR82644	Abt82644 E. coli c
29	109	14.1	23	3	AAB36340	Aab36340
30	109	14.1	23	3	AAB36324	Aab36324
31	109	14.1	23	3	AAB36319	Aab36319
32	104.5	13.5	151	3	AAB36344	Aab36344
33	98.5	12.7	151	3	AAB36342	Aab36342
34	98	12.6	186	6	ABR82645	Abt82645 E. coli c
35	95	12.3	186	6	ABU21488	Abu21488 Protein e
36	93	12.0	19	3	AAB36323	Aab36323
37	93	12.0	19	3	AAB36336	Aab36336
38	93	12.0	19	3	AAB36328	Aab36328
39	92	11.9	23	3	AAB36331	Aab36331
40	92	11.9	673	5	AAV44403	Aay44403 Human tru
41	92	11.9	673	5	AAU79538	Aau79538 Truncated
42	92	11.9	949	5	AAV44404	Aay44404 Human tru
43	92	11.9	949	5	AAU79539	Aau79539 Truncated
44	92	11.9	1327	3	AAV44402	Aay44402 Human can
45	92	11.9	1327	3	AAB27212	Aab27212 Human can

ALIGNMENTS

RESULT 1

ID AAB36347 standard; protein; 151 AA.

XX

AC AAB36347;

XX

DT 26-FEB-2001 (first entry)

XX

DE Agfa::PT3#2 amino acid sequence SEQ ID NO:14.

OS Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;

KW vaccine; immune response; immunogen.

XX

OS Salmonella enteritidis.

OS Escherichia coli.

OS Synthetic.

XX

PN WO200060102-A2.

XX

XX 12-OCT-2000.

XX

PF 05-APR-2000; 2000MO-CA000356.

XX

PR 05-APR-1999; 99US-0127888P.

XX

(UYVI-) UNIV VICTORIA.

XX

PI White AP, Doran JL, Collison SK, Kay WW,

XX

WPI, 2000-672631/65.

DR N-PSDB; AAC64623.

XX

PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence

PT which encodes foreign epitope or antigen, expresses recombinant Agfa

PT protein useful for eliciting immune response in animal.

XX

PS Disclosure; Page 136; 139pp; English.

XX

The present invention describes a recombinant agfa gene (I) where a

segment of the gene has been replaced by a segment of a foreign DNA

sequence which encodes a foreign epitope or antigen. Also described are:

(1) use of thin aggregative fimbriae (SEF17/TRF) nucleation depended

assembly system of strains of Salmonella, Escherichia coli and

CC Enterobacteriaceae for the production of fimbriae comprising recombinant

CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)

CC directing recombination of a recombinant gene into the chromosome of the

homologous species; (3) directing recombination of a recombinant gene

back into the chromosome of the homologous species, replacing the native

PR 05-APR-1999; 99US-0127888P.
XX
XX (UYVI-) UNIV VICTORIA.
XX
XX White AP, Doran JL, Collison SK, Kay WW,
XX WPI; 2000-672631/65.
XX N-PSDB; AAC64631.
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 139; 139pp; English.
XX
XX The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/7AF) nucleation depended
CC assembly system of strains of *Salmonella*, *Escherichia coli* and
CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
CC *Enterobacteriaceae* host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
XX Sequence 151 AA;
SQ
Query Match 85.0%; Score 659; DB 3; Length 151;
Best Local Similarity 81.9%; Pred. No. 9,1e-56;
Matches 136; Conservative 0; Mismatches 0; Indels 30; Gaps 2;
QY 1 MKLLKVAAPFAIIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
DB 1 MKLLKVAAPFAIIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
QY 61 SDAKSETTTTOSGCGAGAVGAGDNGSTTELTONGFRNNATTDONNAKSDITVGGYDQ 120
DB 61 SDAKSETTTTOSGCGAGAVGAGDNGSTTELTONGFRNNATTDONNAKSDITVGGYDQ 120
QY 121 LVTRVVTHEMAHAA-----SYVVRQVFGNNATANQY 151
DB 106 LVTRVVTHEMAHAGNNALVNOTASDSYVVRQVFGNNATANQY 151
RESULT 8
AAB36353
ID AAB36353 standard; protein; 151 AA.
XX
XX AAB36353;
AC
XX 26-FEB-2001 (first entry)
DT
XX
XX Agfa::PT3#8 amino acid sequence SEQ ID NO:26.
DE
XX
XX *Salmonella*, agfa; chromosomal gene replacement; fimbrial; epitope;
KM vaccine; immune response; immunogen.

XX
XX *Salmonella enteritidis*.
OS
XX *Escherichia coli*.
OS
XX Synthetic.
XX
XX W0200060102-A2.
XX
XX 12-OCT-2000.
PD
XX
XX 05-APR-2000; 2000MO-CA000356.
PF
XX
XX 05-APR-1999; 99US-0127888P.
PR
XX
XX (UYVI-) UNIV VICTORIA.
XX
XX White AP, Doran JL, Collison SK, Kay WW,
XX WPI; 2000-672631/65.
XX N-PSDB; AAC64629.
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 138; 139pp; English.
XX
XX The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/7AF) nucleation depended
CC assembly system of strains of *Salmonella*, *Escherichia coli* and
CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
CC *Enterobacteriaceae* host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
XX Sequence 151 AA;
SQ
Query Match 79.0%; Score 612; DB 3; Length 151;
Best Local Similarity 81.5%; Pred. No. 3,2e-51;
Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;
QY 1 MKLLKVAAPFAIIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
DB 1 MKLLKVAAPFAIIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
QY 61 SDAKSETTTTOSGCGAGAVGAGDNGSTTELTONGFRNNATTDONNAKSDITVGGYDQ 120
DB 61 LVTRVVTHEMAHAGYNGAGDNGSTTELTONGFRNNATTDONNAKSDITVGGYDQ 120
QY 121 LVTRVVTHEMAHAAVVRQVFGNNATANQY 151
DB 121 NNAALVNOTASDSYVVRQVFGNNATANQY 151
RESULT 9

AA86349
ID AAB36349 standard; protein; 151 AA.
XX
AC AAB36349;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa::PT3#4 amino acid sequence SEQ ID NO:18.
XX
KM Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
XX vaccine; immune response; immunogen.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO20060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay MW;
XX
DR WPI; 2000-672631/65.
XX N-PSDB; AAC64625.
XX
PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 136; 139pp; English.
XX
XX The present invention describes a recombinant agfa gene (1) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/7AR) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 151 AA;
Query Match 78.8%; Score 611; DB 3; Length 151;
Best Local Similarity 81.5%; Pred. No. 4e-51;
Matches 123; Conservative 6; Mismatches 22; Indels 0; Gaps 0;

61 SDARKSETTITGSGYGADVGAGADNSTLTETLQNGFRNNATIDQWAKNSDITVGQYDQ 120
DB 61 SDARKSETTITGSGYGADVGAGADNSTLTETLQNGFRNNATIDQWAKNSDITVGQYDQ 120
QY 121 LVTRVYTHMAHASYMVRQYFGNNATANY 151
DB 121 NNAALVNOGTASDSSVMVRQYFGNNATANY 151
RESULT 10
ID AAB36350 standard; protein; 151 AA.
XX
AC AAB36350;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa::PT3#5 amino acid sequence SEQ ID NO:20.
XX
KM Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
XX vaccine; immune response; immunogen.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO20060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay MW;
XX
DR WPI; 2000-672631/65.
XX N-PSDB; AAC64626.
XX
PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 137; 139pp; English.
XX
XX The present invention describes a recombinant agfa gene (1) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/7AR) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in

CC the exemplification of the present invention
 XX Sequence 151 AA;
 SQ Query Match 78.6%; Score 609; DB 3; Length 151;
 Best Local Similarity 81.5%; Pred. No. 6.3e-51;
 Matches 122; Conservative 6; Mismatches 22; Indels 0; Gaps 0;

QY 1 MKLLKVAAPFAIIVVSGSALAGVVPQWGGGNNHGGNSGSPDSTLSIYQGSANPAALAQ 60
 DB 1 MKLLKVAAPFAIIVVSGSALAGVVPQWGGGNNHGGNSGSPDSTLSIYQGSANPAALAQ 60
 QY 61 SDARKSETTTTQSGYGNAGADVGGADNSTIELTQNGFRNNATTDQNNAKNSDITVGGYDQ 120
 DB 61 SDARKYDQLVTRVVTHEMAHAHVAGGADNSTIELTQNGFRNNATTDQNNAKNSDITVGGYDQ 120
 QY 121 LVTRVVTHEMAHSAVWVROYGFGNNATANQY 151
 DB 121 NNAAALVNTASDSSVMVROYGFGNNATANQY 151

RESULT 11
 AAB36354
 ID AAB36354 standard; protein; 151 AA.
 AC AAB36354;
 XX 26-FEB-2001 (first entry)
 DT Agfa::PT3#9 amino acid sequence SEQ ID NO:28.
 DE Agfa::PT3#9 amino acid sequence SEQ ID NO:28.
 XX Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 XX Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX WO200060102-A2.
 PN 12-OCT-2000.
 PD 05-APR-2000; 2000WO-CA000356.
 PF 05-APR-1999; 99US-0127888P.
 PR (UYVT-) UNIV VICTORIA.
 PA White AP, Doran JL, Collison SK, Kay WW;
 PI WPI; 2000-672631/65.
 DR N-PSDB; AAC64630.
 XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX Disclosure; Page 138; 139pp; English.

CC The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal.
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of a foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or

CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

SQ Sequence 151 AA;
 QY Query Match 77.8%; Score 603; DB 3; Length 151;
 Best Local Similarity 80.8%; Pred. No. 2.4e-50;
 Matches 122; Conservative 4; Mismatches 25; Indels 0; Gaps 0;

QY 1 MKLLKVAAPFAIIVVSGSALAGVVPQWGGGNNHGGNSGSPDSTLSIYQGSANPAALAQ 60
 DB 1 MKLLKVAAPFAIIVVSGSALAGVVPQWGGGNNHGGNSGSPDSTLSIYQGSANPAALAQ 60
 QY 61 SDARKSETTTTQSGYGNAGADVGGADNSTIELTQNGFRNNATTDQNNAKNSDITVGGYDQ 120
 DB 61 SDARKSETTTTQSGYGNAGADVGGADNSTIELTQNGFRNNATTDQNNAKNSDITVGGYDQ 120
 QY 121 LVTRVVTHEMAHSAVWVROYGFGNNATANQY 151
 DB 121 NNAAALVNTASDSSVMVROYGFGNNATANQY 151

RESULT 12
 AAB36351
 ID AAB36351 standard; protein; 151 AA.
 AC AAB36351;
 XX 26-FEB-2001 (first entry)
 DT Agfa::PT3#6 amino acid sequence SEQ ID NO:22.
 DE Agfa::PT3#6 amino acid sequence SEQ ID NO:22.
 XX Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 XX Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX WO200060102-A2.
 PN 12-OCT-2000.
 PD 05-APR-2000; 2000WO-CA000356.
 PF 05-APR-1999; 99US-0127888P.
 PR (UYVT-) UNIV VICTORIA.
 PA White AP, Doran JL, Collison SK, Kay WW;
 PI WPI; 2000-672631/65.
 DR N-PSDB; AAC64627.
 XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX Disclosure; Page 137; 139pp; English.

CC The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:25:54 ; Search time 12.4 Seconds
(without alignments)
628,671 Million cell updates/sec

Title: US-09-543-407-14

Perfect score: 775

Sequence: 1 MKLLKVAFAAIVVSGSALA.....HASVMYQVGFNNATANDY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA: *
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PTUS.COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfl1es1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	692	89.3	151	1 US-08-233-788A-59	Sequence 59, Appl
2	514	66.3	120	1 US-08-233-788A-57	Sequence 57, Appl
3	92	11.9	673	1 US-09-196-387-8	Sequence 8, Appl
4	92	11.9	673	4 US-09-841-835-8	Sequence 8, Appl
5	92	11.9	949	4 US-09-196-387-10	Sequence 10, Appl
6	92	11.9	949	4 US-09-841-835-10	Sequence 10, Appl
7	92	11.9	1327	3 US-09-196-387-2	Sequence 2, Appl
8	92	11.9	1327	4 US-09-841-835-2	Sequence 2, Appl
9	92	11.9	1327	4 US-09-972-115A-8	Sequence 8, Appl
10	89.5	11.5	738	3 US-08-864-038A-3	Sequence 3, Appl
11	87	11.2	2123	3 US-08-968-685A-10	Sequence 10, Appl
12	83.5	10.8	3139	4 US-09-252-991A-32096	Sequence 32096, A
13	83.5	10.8	943	4 US-09-056-556-204	Sequence 204, App
14	83.5	10.8	943	4 US-09-072-596-199	Sequence 199, App
15	83.5	10.8	943	4 US-09-477-135A-131	Sequence 131, App
16	83.5	10.8	943	4 US-09-072-967-204	Sequence 204, App
17	82.5	10.6	2736	4 US-09-252-991A-30227	Sequence 30227, A
18	81	10.5	878	4 US-09-540-336-3401	Sequence 3401, Ap
19	80.5	10.4	873	4 US-09-336-447A-13	Sequence 13, Appl
20	80	10.3	892	4 US-09-336-447A-5	Sequence 5, Appl
21	79.5	10.3	204	6 5187262-1	Patent No. 5187262
22	79.5	10.3	518	3 US-09-043-123-2	Sequence 123, A
23	79.5	10.3	1207	4 US-09-489-039A-11518	Sequence 11518, A
24	78.5	10.1	539	4 US-09-719-402A-2	Sequence 2, Appl
25	78.5	10.1	714	4 US-09-841-786-4	Sequence 4, Appl
26	78.5	10.1	1912	1 US-08-409-995-4	Sequence 4, Appl
27	78.5	10.1	1912	3 US-08-685-467-4	Sequence 4, Appl

28	78.5	10.1	2353	3 US-09-377-155-33	Sequence 33, Appl
29	78.5	10.1	2353	3 US-08-913-942-4	Sequence 4, Appl
30	78.5	10.1	2353	4 US-09-669-974-33	Sequence 33, Appl
31	78.5	10.1	2353	4 US-09-797-862-33	Sequence 33, Appl
32	78.5	10.1	2354	4 US-09-268-347-47	Sequence 47, Appl
33	78.5	10.1	2411	4 US-09-268-347-36	Sequence 36, Appl
34	78.5	10.1	3241	4 US-09-841-786-1	Sequence 1, Appl
35	77.5	10.0	318	4 US-09-107-532A-5818	Sequence 5818, Ap
36	77.5	10.0	3169	4 US-09-453-702B-257	Sequence 257, App
37	77	9.9	415	4 US-09-025-769B-280	Sequence 280, App
38	77	9.9	1690	4 US-09-595-684B-39	Sequence 39, Appl
39	76.5	9.9	956	4 US-09-495-880A-42	Sequence 42, Appl
40	76.5	9.9	956	4 US-09-134-078-63	Sequence 63, Appl
41	76.5	9.9	1413	4 US-09-252-991A-21268	Sequence 21268, A
42	75.5	9.7	645	4 US-09-919-172-41	Sequence 41, Appl
43	75.5	9.7	789	3 US-08-960-780-6	Sequence 6, Appl
44	75.5	9.7	789	3 US-09-073-898-6	Sequence 6, Appl
45	75.5	9.7	789	4 US-09-850-351A-6	Sequence 6, Appl

ALIGNMENTS

RESULT 1
US-08-233-788A-59
; Sequence 59, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, Karen S.
; APPLICANT: Clouthier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; TITLE OF INVENTION: OF SALMONELLA
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,788A
; FILING DATE: 26-Apr-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEDANBERY
; INFORMATION FOR SEQ ID NO: 59:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 151 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-233-788A-59
Query Match 89.3%; Score 692; DB 1; Length 151;
Best Local Similarity 90.7%; Pred. No. 3.9e-62;
Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;
1 MKLLKVAFAAIVVSGSALAGVTPQWGGGNNHNGGSSGPDSTLSTYQGSNMAALALQ 60
|||||

Db 1 MKLKVAFAAIVVSGSALAGVVPQWGGGNNHGGSSGPDSTLSIYOGSANAALALD 60
QY 61 SDARKSETTITQSGYGVGADVGQADNSTIELTONGFRNNATIDOMNAKSDITVGOYDQ 120
Db 61 SDARKSETTITQSGYGVGADVGQADNSTIELTONGFRNNATIDOMNAKSDITVGOYDQ 120
QY 121 LVTRVTHEMAHSAVMVROVGFQGNNTATANOY 151
Db 121 NNPALVNGTASDSSVWVRQVGFQGNNTATANOY 151

RESULT 2
US-08-233-788A-57

Sequence 57, Application US/08233788A
Patent No. 5635617

GENERAL INFORMATION:

APPLICANT: Doran, James L.
APPLICANT: Kay, William W.
APPLICANT: Collinson, Karen S.
APPLICANT: Clouthier, Sharon C.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
TITLE OF INVENTION: OF SALMONELLA
NUMBER OF SEQUENCES: 61
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed and Berry
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: U.S.A.
ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233,788A
FILING DATE: 26-APR-1994
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: King, Joshua
REGISTRATION NUMBER: 35,570
REFERENCE/DOCKET NUMBER: 920043.403C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031

TELEX: 3723836 SEDANBERRY

INFORMATION FOR SEQ ID NO: 57:

SEQUENCE CHARACTERISTICS:

LENGTH: 120 amino acids
TYPE: amino acid
TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-233-788A-57

Query Match 66.3%; Score 514; DB 1; Length 120;
Best Local Similarity 100.0%; Pred. No. 2.2e-44;

Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 VVPWGGGNNHGGSSGPDSTLSIYOGSANAALALQSDARKSETTITQSGYGVGADV 81
Db 1 VVPWGGGNNHGGSSGPDSTLSIYOGSANAALALQSDARKSETTITQSGYGVGADV 60
QY 82 GQAGDNSTIELTONGFRNNATIDOMNAKSDITVGOY 118
Db 61 GQAGDNSTIELTONGFRNNATIDOMNAKSDITVGOY 97

RESULT 3

US-09-196-387-8

Sequence 8, Application US/09196387

Patent No. 6277613

GENERAL INFORMATION:

APPLICANT: de Lange, Titia

APPLICANT: Smith, Susan

TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS

TITLE OF INVENTION: OF USE THEREOF

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSEE: Klauber & Jackson

STREET: 411 Hackensack Avenue, 4th floor

CITY: Hackensack

STATE: New Jersey

COUNTRY: USA

ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/196,387

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/095,225

FILING DATE: June 10, 1998

ATTORNEY/AGENT INFORMATION:

NAME: Jackson Esq., David A.

REGISTRATION NUMBER: 26,742

REFERENCE/DOCKET NUMBER: 600-1-230 CIP1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201-487-5800

TELEFAX: 201-343-1684

TELEX: 133521

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 673 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-196-387-8

Query Match 11.9%; Score 92; DB 3; Length 673;

Best Local Similarity 30.4%; Pred. No. 0.54;

Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

QY 6 VAAFAAI-VVSGSALAGVVPQWGGGNNHGGSSGPDSTLSIYOGSANAALALQSDAR 64
Db 99 VAAAPVPAVSTSSAAGVAPNPAAGSGNNPSSSSPTSS--SSSPSSPGSSLAESPAA 157
QY 65 KSETTIT---OSGYGVGADVGQADNSTIELTONG--FRNNATIDOMNAKSDI 113
Db 158 GVSSTAPLGPAAAGPOTGVPAVSGALRELLACRNGDVSRVRLVDANVANAKDM 212

RESULT 4

US-09-841-835-8

Sequence 8, Application US/09841835

Patent No. 6506587

GENERAL INFORMATION:

APPLICANT: de Lange, Titia

TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS

TITLE OF INVENTION: OF USE THEREOF

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSEE: Klauber & Jackson

STREET: 411 Hackensack Avenue, 4th floor

CITY: Hackensack

STATE: New Jersey

COUNTRY: USA

ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

Db 99 VAAAPVPAVSTSSAGVAPNPAGSGNNSSPSSSSPTSS-SSSSPSSPSSSLAESPEEA 157
 QY 65 KSETTIT---QSGYNGADVQGGADNSTIELTONG--FRNNATIDQWNAKNSDI 113
 Db 158 GVSSTAPLPGAGAPGTGTGVPVAVSGALRELLACRNGDVSRVKLVDAANVNAKOM 212

RESULT 7

US-09-196-387-2
 ; Sequence 2, Application US/09196387
 ; Patent No. 6277613
 ; GENERAL INFORMATION:

APPLICANT: de Lange, Titia
 APPLICANT: Smith, Susan
 TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
 TITLE OF INVENTION: OF USE THEREOF
 NUMBER OF SEQUENCES: 12
 CORRESPONDENCE ADDRESS:

ADDRESSEE: Klauber & Jackson
 STREET: 411 Hackensack Avenue, 4th Floor
 CITY: Hackensack
 STATE: New Jersey
 COUNTRY: USA
 ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Releasee #1.0, Version #1.30
 CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/196,387
 FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 09/095,225
 FILING DATE: June 10, 1998

ATTORNEY/AGENT INFORMATION:

NAME: Jackson Esq., David A.
 REGISTRATION NUMBER: 26,742
 REFERENCE/DOCKET NUMBER: 600-1-230 CIP1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201-487-5800
 TELEFAX: 201-343-1684
 TELEX: 133521

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1327 amino acids
 TYPE: amino acid
 STRANDEDNESS: single

TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHETICAL: NO

US-09-196-387-2

Query Match 11.9%; Score 92; DB 3; Length 1327;
 Best Local Similarity 30.4%; Pred. No. 1.3; Mismatches 57; Indels 8; Gaps 4;

Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;
 QY 6 VAAAPAI-VVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANAALALQSDAR 64
 Db 99 VAAAPVPAVSTSSAGVAPNPAGSGNNSSPSSSSPTSS-SSSSPSSPSSSLAESPEEA 157
 QY 65 KSETTIT---QSGYNGADVQGGADNSTIELTONG--FRNNATIDQWNAKNSDI 113
 Db 158 GVSSTAPLPGAGAPGTGTGVPVAVSGALRELLACRNGDVSRVKLVDAANVNAKOM 212

RESULT 8

US-09-841-835-2

; Sequence 2, Application US/09841835
 ; Patent No. 6506587
 ; GENERAL INFORMATION:

APPLICANT: de Lange, Titia
 APPLICANT: Smith, Susan
 TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
 TITLE OF INVENTION: OF USE THEREOF
 NUMBER OF SEQUENCES: 12
 CORRESPONDENCE ADDRESS:

ADDRESSEE: Klauber & Jackson
 STREET: 411 Hackensack Avenue, 4th Floor
 CITY: Hackensack
 STATE: New Jersey
 COUNTRY: USA
 ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Releasee #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/841,835
 FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 09/196,387
 FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Jackson Esq., David A.
 REGISTRATION NUMBER: 26,742
 REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 201-487-5800
 TELEFAX: 201-343-1684
 TELEX: 133521

INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1327 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHETICAL: NO

US-09-841-835-2

Query Match 11.9%; Score 92; DB 4; Length 1327;
 Best Local Similarity 30.4%; Pred. No. 1.3; Mismatches 57; Indels 8; Gaps 4;

Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;
 QY 6 VAAAPAI-VVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANAALALQSDAR 64
 Db 99 VAAAPVPAVSTSSAGVAPNPAGSGNNSSPSSSSPTSS-SSSSPSSPSSSLAESPEEA 157
 QY 65 KSETTIT---QSGYNGADVQGGADNSTIELTONG--FRNNATIDQWNAKNSDI 113
 Db 158 GVSSTAPLPGAGAPGTGTGVPVAVSGALRELLACRNGDVSRVKLVDAANVNAKOM 212

RESULT 9

US-09-972-115A-8

; Sequence 8, Application US/09972115A
 ; Patent No. 6599728
 ; GENERAL INFORMATION:

APPLICANT: Geron Corporation
 APPLICANT: Gregg, Martin B.
 APPLICANT: Walter, Funk D.
 APPLICANT: Mieczyslaw, Piatyszek A.

TITLE OF INVENTION: A Second Mammalian Telomerase
 FILE REFERENCE: 080/003C

CURRENT APPLICATION NUMBER: US/09/972,115A
 PRIOR FILING DATE: 2001-10-05
 PRIOR APPLICATION NUMBER: US 60/128,577

PRIOR FILING DATE: 2000-04-10
 PRIOR APPLICATION NUMBER: US 60/129,123
 PRIOR FILING DATE: 1999-04-13

RESULT 10

US-09-841-835-2

; Sequence 2, Application US/09841835
 ; Patent No. 6506587
 ; GENERAL INFORMATION:

APPLICANT: de Lange, Titia
 APPLICANT: Smith, Susan
 TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
 TITLE OF INVENTION: OF USE THEREOF
 NUMBER OF SEQUENCES: 12
 CORRESPONDENCE ADDRESS:

SOFTWARE: PatentIn version 3.1
SEQ ID NO: 8
LENGTH: 1327
TYPE: PRT
ORGANISM: Homo sapiens
US-09-972-115A-8

Query Match 11.9%; Score 92; DB 4; Length 1327;
Best Local Similarity 30.4%; Pred. No. 1.3;
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

QY 6 VAAFAAI-VVSGALGVVPOMGGGHHNGGSSGPDSTLSIYGSANALALQSDAR 64
DB 99 VAAAPVPAVSTSSAGVAPNPAGSGSNSSSSSTSS-SSSSPSSSLAESPEAA 157
QY 65 KSETTIT---OSGYNGADVGGAGADNSTELTONG--FNNNATIDQWNAKNSDI 113
DB 158 GVSTNAPLGRGMAOPGTGVPAVSGALRELEACRNGVSRVXKLVDAANVNAADM 212

RESULT 10
US-08-864-038A-3

Sequence 3, Application US/08864038A
Patent No. 6001592

GENERAL INFORMATION:

APPLICANT: KUNIO NAKASHIMA et al.

TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE CDNA, VECTOR

TITLE OF INVENTION: CONTAINING SAID CDNA, HOST CELLS TRANSFORMED WITH SAID

TITLE OF INVENTION: VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING

TITLE OF INVENTION: SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBODY

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESS: 812-5 Hirano

STREET: Ieshinden

CITY: Ten-city

STATE: Mie-prefecture

COUNTRY: JAPAN

ZIP: 514-01

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage

COMPUTER: IBM Compatible

OPERATING SYSTEM: Microsoft Windows 95

SOFTWARE: Word Perfect 6.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/864,038A

FILING DATE: May 28, 1997

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 8-184459

FILING DATE: 15-July-1996

ATTORNEY/AGENT INFORMATION:

NAME: C. Bruce Hamburg

REGISTRATION NUMBER: 22,389

REFERENCE/DOCKET NUMBER: F-5610

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212)986-2340

TELEFAX: (212)953-7733

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 738

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

ORIGINAL SOURCE:

ORGANISM: Pinctada fucata

CELL TYPE: mantle epithelial cell

FEATURE:

NAME/KEY: peptide

LOCATION: from 1 to 738

IDENTIFICATION METHOD: E (by experiment)

US-08-864-038A-3

Query Match 11.5%; Score 89.5; DB 3; Length 738;

Best Local Similarity 35.4%; Pred. No. 1.1;
Matches 29; Conservative 4; Mismatches 26; Indels 23; Gaps 3;

QY 413 LKVAAPAIIVVSGALGVVPOMGGGHHNGGSSGPDSTLSIYGSANALALQSDAR 62
DB 419 LKSSASASASASASASAG-----GGGGGGGGGGGGG-----GGAGALAA----- 460

QY 63 ARKSETTITOSGYNGADVGGG 84
DB 461 -----AALAAAGAGGGLGGGGG 477

RESULT 11

US-08-968-685A-10

Sequence 10, Application US/08968685A

Patent No. 6214981

GENERAL INFORMATION:

APPLICANT: TUCKER, KENNETH

TITLE OF INVENTION: PLOSILA, LAURA

TITLE OF INVENTION: MORAXELLA CATARRHALIS OUTER MEMBRANE

TITLE OF INVENTION: PROTEIN-106 POLYPEPTIDE, GENE SEQUENCE AND USES THEREOF

NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESS:

ADDRESS: PENNIE & EDMONDS LLP

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/968,685A

FILING DATE: No. 6214981ember 12, 1997

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Baldwin, Geraldine F.

REGISTRATION NUMBER: 31,232

REFERENCE/DOCKET NUMBER: 7969-060

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 790-9090

TELEFAX: (212) 869-8864

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 2123 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: unknown

MOLECULE TYPE: peptide

US-08-968-685A-10

Query Match 11.2%; Score 87; DB 3; Length 2123;
Best Local Similarity 25.4%; Pred. No. 7.6;
Matches 48; Conservative 27; Mismatches 62; Indels 52; Gaps 11;

QY 1 MKLKVAPPAIVV-----SGSALAGVVPOMGGGHHNGGSSGPDSTLSIYGSANAA 56
DB 44 LSPARIALAVLVGTATNGSAYAGIGISEADGG--KGGANAG--DKSIAI---GDAQA 97

QY 57 LMSDARKSETTITOSGYNGADV--OGADNSTI-----ELTONGFRNNA--TIQWNAK 109
DB 98 LGSQSLAIGDNKIVANS--NNNANTGAKASGNEISAIIGDVLASGASIALISGDDLYLK 155

QY 110 NSDITVGYDQDLVTRV-----VTHEMAASVMYRQV----- 140
DB 156 E---TVQQLSELLPIIRGGKALNDIVQLADTNLQKTRRTHAQGHASTAVGAMSYAKGHFS 212
QY 141 -GRGNATA 148

Db 213 NAFGRATA 221

RESULT 12

US-09-252-991A-32096
Sequence 32096, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
PRIOR FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 32096
LENGTH: 339
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-32096

Query Match 10.8%; Score 83.5; DB 4; Length 339;
Best Local Similarity 21.5%; Pred. No. 1.6;
Matches 35; Conservative 28; Mismatches 53; Indels 47; Gaps 7;

QY 5 KVAAPAAIVVSSALAGVPWQNG-----GGNNGGSSGPDSTLIYQGSANAAL 57
DB 58 KVSNGTGL--NNASVSGSIKASGVVAVVAGDNNQANMAAALASADASFVFGATAS- 114
QY 58 ALQSDARKSETTITQSGYNGGADVGQADNNTIELTONGFRNNATTIDQWAKNS----- 111
DB 115 -----TSVLOSQVGN-----TLNNISNPTASISNSANNVSGNLG 149
QY 112 -DITVGYQDQ---LVTRVVTHEMAASVMVRQVFGNNATPAN 149
DB 150 VVVAAGNFQGNKDILAAVAVNSQYSTAGSAASQSTG-NTTVN 191

RESULT 13

US-09-056-556-204
Sequence 204, Application US/09056556
Patent No. 6350456
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
NUMBER OF SEQUENCES: 241
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/056,556
FILING DATE: 07-APR-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.457
TELECOMMUNICATION INFORMATION:

TREATY

TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 204:
SEQUENCE CHARACTERISTICS:
LENGTH: 943 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-09-056-556-204

Query Match 10.8%; Score 83.5; DB 4; Length 943;
Best Local Similarity 25.4%; Pred. No. 6;
Matches 29; Conservative 16; Mismatches 50; Indels 19; Gaps 4;

QY 12 IVVSGSLAIVPWQGGGNGHNGGSSGPDSTLIYQGSANAALALQSDAR--KSET 68
DB 159 IGLTSGSLGTF-----GGINSQTN-----IGLFNSGTGVVIGNSGTGMJGIGNG 205
QY 69 TITQSGYNGADVGQADNNTIELTONGFRNNATTIDQWAKNSD---ITVGYQD 119
DB 206 NSVNTGFGNSGDANTGFGNSGIANVGAGNNTGSGYNPNSNTGCFNNQGYN 259

RESULT 14

US-09-072-596-199
Sequence 199, Application US/09072596
Patent No. 6458366
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Nero, Antonia
APPLICANT: Houghton, Raymond
APPLICANT: Vedrick, Thomas S.
APPLICANT: Twardzik, Daniel R.
APPLICANT: Lodes, Michael J.
APPLICANT: Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
NUMBER OF SEQUENCES: 350
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/072,596
FILING DATE: 05-MAY-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 199:
SEQUENCE CHARACTERISTICS:
LENGTH: 943 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-09-072-596-199

Query Match 10.8%; Score 83.5; DB 4; Length 943;
Best Local Similarity 25.4%; Pred. No. 6;
Matches 29; Conservative 16; Mismatches 50; Indels 19; Gaps 4;

[illegible]

RESULT 15
TIC-09-477

```

Sequence 131. Application US/09477135A
Patent No. 6572865
GENERAL INFORMATION:
APPLICANT: Nano, Francis
TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
FILE OF INVENTION: Immunostimulatory Peptides
FILE REFERENCE: 52868
CURRENT APPLICATION NUMBER: US/09/477,135A
CURRENT FILING DATE: 2000-01-03
PRIOR APPLICATION NUMBER: 08990823
PRIOR FILING DATE: 1997-12-15
PRIOR APPLICATION NUMBER: US 96/10375
PRIOR FILING DATE: 1996-06-14
PRIOR APPLICATION NUMBER: 60/000,254
PRIOR FILING DATE: 1995-06-15
NUMBER OF SEQ ID NOS: 169
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 131
LENGTH: 943
TYPE: PRT
ORGANISM: Mycobacterium tuberculosis
US-09-477-135A-131

```

Query Match	10.8%	Score 83.5	DB 4	Length 943
Best Local Similarity	25.4%	Pred. No. 6		
Matches 29	Conservative 16	Mismatches 50	Indels 19	Gaps 4

```

QY      12  IVSGSLGAGVVPQMGSGGNHNGGSGSPDSTLSLYQGASNALALAQSDAR---KSET 68
Db      266  IGLTGSGLG-----GGINSTGN-----IGLFGSGTGNVIGISGRTGNMGDINS 312
QY      69  TTITQSGVGNADVGGAGDNSTIELTONGPRNNATIDOMAKSD---ITVQYD 119
Db      313  NSVNTGNSGNDANTPFNSGILANGVGAAGVNTGTSVTPGNSNNGGFMNQYN 366

```

Search completed: March 11, 2004, 18:44:45
Job time : 34.4 secs

THIS PAGE BLANK (USPTO)

Db 78 GNSNSV-----GRDIQKQSGAGNSAIFQEGTGSVDELOQTGTSNGAVPSGMMWN 129
Qy 111 SDITVGOYDQVTVTHVTEHMAHAYM-----VROVFGNATANO 150
Db 130 DP---GVFNK-ITQDSSNSGKSVIODKNNVFSIKQGTGNTGSTVNO 174

RESULT 2

US-09-841-835-8
; Sequence 8, Application US/09841835
; Patent No. US20020076795A1
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; TITLE OF INVENTION: OF USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/841,835
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/196,387
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 673 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-841-835-8

Query Match 11.9%; Score 92; DB 9; Length 673;
Best Local Similarity 30.4%; Pred. No. 0.85;

Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

Qy 6 VAAFAAI-VVSGSALAGVVPQWGGGNNHNGGSSGPPSTLSTIYQGSANALALQSDAR 64
Db 99 VAAAVVPAVSTSSAAGVAPNPAGSGSNSSPSSPTSS-SSSSPSSPGSLAESPEEA 157
Qy 65 KSETTIT---QSGYGNADVGQAGDNSTIELTONG--FRNNATTIDQNNAKNSDI 113
Db 158 GVSTAPLPGPAGAGGTGVPVAVSGALRELLACRNGDVSVKRLVDAAVNAKDM 212

RESULT 3

US-09-841-835-10
; Sequence 10, Application US/09841835
; Patent No. US20020076795A1
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS

; TITLE OF INVENTION: OF USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/841,835
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/196,387
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 949 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-841-835-10

Query Match 11.9%; Score 92; DB 9; Length 949;
Best Local Similarity 30.4%; Pred. No. 1.3;

Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

Qy 6 VAAFAAI-VVSGSALAGVVPQWGGGNNHNGGSSGPPSTLSTIYQGSANALALQSDAR 64
Db 99 VAAAVVPAVSTSSAAGVAPNPAGSGSNSSPSSPTSS-SSSSPSSPGSLAESPEEA 157
Qy 65 KSETTIT---QSGYGNADVGQAGDNSTIELTONG--FRNNATTIDQNNAKNSDI 113
Db 158 GVSTAPLPGPAGAGGTGVPVAVSGALRELLACRNGDVSVKRLVDAAVNAKDM 212

RESULT 4

US-09-841-835-2
; Sequence 2, Application US/09841835
; Patent No. US20020076795A1
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; TITLE OF INVENTION: OF USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/841,835
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/196,387
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1327 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-09-841-835-2

Query Match 11.9%; Score 92; DB 9; Length 1327;
Best Local Similarity 30.4%; Pred. No. 2;
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

QY 6 VAAFAAI-VVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQSDAR 64
DB 99 VAAAPVPAVSTSSAAGVAPNPAGSGSNNSSPSSSSPTSS-SSSSPSSPSSSLAESPEAA 157
QY 65 KSETTT---OSGYNGADVGGADNSTIELTQNG--FRNNATTIQMAKNSDI 113
DB 158 GVSSTAPLPGGAAGPGTGVPVAVSGALRELLACRNGDVSRVKRLVDAANVAKDM 212

RESULT 5
US-09-972-115A-8
Sequence 8; Application US/09972115A
Publication No. US20030032769A1
GENERAL INFORMATION:
APPLICANT: Geron Corporation
APPLICANT: Gregg, Morin B.
APPLICANT: Walter, Funk D.
APPLICANT: Mieczyslaw, Piatyzek A.
TITLE OF INVENTION: A Second Mammalian Telomerase
FILE REFERENCE: 080/003C
CURRENT APPLICATION NUMBER: US/09/972,115A
CURRENT FILING DATE: 2001-10-05
PRIOR APPLICATION NUMBER: US 60/128,577
PRIOR FILING DATE: 2000-04-10
PRIOR APPLICATION NUMBER: US 60/129,123
PRIOR FILING DATE: 1999-04-13
NUMBER OF SEQ ID NOS: 64
SOFTWARE: Patentin version 3.1
SEQ ID NO 8
LENGTH: 1327
TYPE: PRT
ORGANISM: Homo sapiens
US-09-972-115A-8

Query Match 11.9%; Score 92; DB 10; Length 1327;
Best Local Similarity 30.4%; Pred. No. 2;
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

QY 6 VAAFAAI-VVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQSDAR 64
DB 99 VAAAPVPAVSTSSAAGVAPNPAGSGSNNSSPSSSSPTSS-SSSSPSSPSSSLAESPEAA 157
QY 65 KSETTT---OSGYNGADVGGADNSTIELTQNG--FRNNATTIQMAKNSDI 113
DB 158 GVSSTAPLPGGAAGPGTGVPVAVSGALRELLACRNGDVSRVKRLVDAANVAKDM 212

RESULT 6
US-10-199-937-4
Sequence 4; Application US/10199937
Publication No. US20030190739A1
GENERAL INFORMATION:
APPLICANT: Christenson, Erik
APPLICANT: Demaggio, Anthony J.
APPLICANT: Goldman, Phyllis S.
APPLICANT: McElligott, David L.
TITLE OF INVENTION: TANKRASE2 MATERIALS AND METHODS
FILE REFERENCE: 27866/36559
CURRENT APPLICATION NUMBER: US/10/199,937
CURRENT FILING DATE: 2002-07-22
PRIOR APPLICATION NUMBER: US/09/606,035
PRIOR FILING DATE: 2000-06-28
PRIOR APPLICATION NUMBER: 60/141,582
NUMBER OF SEQ ID NOS: 178
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 4
LENGTH: 1327
TYPE: PRT
ORGANISM: Homo sapiens
US-10-199-937-4

Query Match 11.9%; Score 92; DB 14; Length 1327;
Best Local Similarity 30.4%; Pred. No. 2;
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

QY 6 VAAFAAI-VVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQSDAR 64
DB 99 VAAAPVPAVSTSSAAGVAPNPAGSGSNNSSPSSSSPTSS-SSSSPSSPSSSLAESPEAA 157
QY 65 KSETTT---OSGYNGADVGGADNSTIELTQNG--FRNNATTIQMAKNSDI 113
DB 158 GVSSTAPLPGGAAGPGTGVPVAVSGALRELLACRNGDVSRVKRLVDAANVAKDM 212

RESULT 7
US-09-793-306-146
Sequence 146; Application US/09793306
Patent No. US20020098200A1
GENERAL INFORMATION:
APPLICANT: Campos-Neto, Antonio
APPLICANT: Skelky, Yasir
APPLICANT: Ovendale, Pamela
APPLICANT: Jen, Shyian
APPLICANT: Lodes, Michael
APPLICANT: Corixa Corporation
TITLE OF INVENTION: Compounds and Methods for Diagnosis and Immunotherapy
FILE REFERENCE: 014058-008740US
CURRENT APPLICATION NUMBER: US/09/793,306
CURRENT FILING DATE: 2001-02-26
PRIOR APPLICATION NUMBER: US 60/185,037
PRIOR FILING DATE: 2000-02-25
PRIOR APPLICATION NUMBER: US 60/223,828
NUMBER OF SEQ ID NOS: 164
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 146
LENGTH: 597
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: mTTC#3-His
US-09-793-306-146

Query Match 11.2%; Score 87; DB 9; Length 597;
Best Local Similarity 26.6%; Pred. No. 2.4;
Matches 33; Conservative 15; Mismatches 50; Indels 26; Gaps 5;

QY 26 MCGGNNNGGNSGSDSTLSTIYQGSANALALQSDARKSETTITQSGYNGADVQGA 85
 Db 358 FGNSGNNNIGFNSG--NNNVGFNSGNNNFGNAGD-----INTGFNAGDINTGF 408
 QY 86 DNSTIELTQNGFRNNATIDQNAKNSDITVQDQVTVVTHEMAHSAVMRQVFGNN 145
 Db 409 GNA-----GFEN--MGIGNAGNEDVGVNGSGSTNVGV-----GNAQNSVGFQNA 451
 QY 146 ATAN 149
 Db 452 GTLN 455

RESULT 8

US-09-813-214A-9
 ; Sequence 9, Application US/09813214A
 ; Patent No. US20020177200A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Tucker, Kenneth
 ; APPLICANT: Piosella, Laura
 ; TITLE OF INVENTION: MORAXELLA CATARRHALIS OUTER MEMBRANE PROTEIN-106 POLYPEPTIDE, GEN
 ; FILE REFERENCE: 7969-089-999
 ; CURRENT APPLICATION NUMBER: US/09/813,214A
 ; CURRENT FILING DATE: 2000-03-20
 ; PRIOR APPLICATION NUMBER: 08/968,685
 ; PRIOR FILING DATE: 1997-11-12
 ; NUMBER OF SEQ ID NOS: 20
 ; SOFTWARE: PatentIn Version 3.0
 ; SEQ ID NO 9
 ; LENGTH: 2122
 ; TYPE: PRT
 ; ORGANISM: Moraxella catarrhalis
 US-09-813-214A-9

Query Match 11.2%; Score 87; DB 9; Length 2122;
 Best Local Similarity 25.4%; Pred. No. 12;
 Matches 48; Conservative 27; Mismatches 62; Indels 52; Gaps 11;

QY 1 MGLLVAAFAAIVV-----SGSALAGVNVQWGGGNNHGGNSGSDSTLSTIYQGSANAA 56
 Db 44 LSPFAIALLAVIVATNTANGSNVYGIIGSEADGG--KGGANARG-DKSIAL---GDIQAQ 97
 QY 57 LALQSDARKSETTITQSGYNGADVQ--QGANSTI---ELTQNGFRNNA--TTIDQNAK 109
 Db 98 LGSQGIALGDMKIVHNS--NNNANIGAKASGNSISALGSDVLASGHASIALGSDDLVKK 155
 QY 110 NSDITVQGYDQVTVRV-----VTHEMAHSAVMRQV----- 140
 Db 156 E---TVQGISLPLTIRGOKALNDIYQADTNLQKRYRTAOGHASTAVGMSYAKGHFS 212
 QY 141 -GFGNNATA 148
 Db 213 NAFGRATA 221

RESULT 9

US-09-712-363-156
 ; Sequence 156, Application US/09712363
 ; Patent No. US20020164588A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Eisenberg, David
 ; APPLICANT: Rosenstein, Sergio H.
 ; APPLICANT: Marcotte, Edward M.
 ; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
 ; FILE REFERENCE: 07419-032001
 ; CURRENT APPLICATION NUMBER: US/09/712,363
 ; CURRENT FILING DATE: 2000-11-13
 ; PRIOR APPLICATION NUMBER: PCT/US00/02246
 ; PRIOR FILING DATE: 2000-01-28
 ; PRIOR APPLICATION NUMBER: 60/179,531

; PRIOR FILING DATE: 2000-02-01
 ; PRIOR APPLICATION NUMBER: 60/117,844
 ; PRIOR FILING DATE: 1999-01-29
 ; PRIOR APPLICATION NUMBER: 60/118,206,
 ; PRIOR FILING DATE: 1999-02-01
 ; PRIOR APPLICATION NUMBER: 60/126,593
 ; PRIOR FILING DATE: 1999-03-26
 ; PRIOR APPLICATION NUMBER: 60/134,093
 ; PRIOR FILING DATE: 1999-05-14
 ; PRIOR APPLICATION NUMBER: 60/134,092
 ; PRIOR FILING DATE: 1999-05-14
 ; PRIOR APPLICATION NUMBER: 60/165,124
 ; PRIOR FILING DATE: 1999-11-12
 ; PRIOR APPLICATION NUMBER: 60/165,086
 ; PRIOR FILING DATE: 1999-11-12
 ; NUMBER OF SEQ ID NOS: 292
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 156
 ; LENGTH: 527
 ; TYPE: PRT
 ; ORGANISM: Mycobacterium tuberculosis
 US-09-712-363-156

Query Match 11.2%; Score 86.5; DB 9; Length 527;
 Best Local Similarity 22.9%; Pred. No. 2.3;
 Matches 35; Conservative 13; Mismatches 56; Indels 49; Gaps 5;

QY 21 GVVPQWGG---GGNNHGGNSGSD-----STLSIYQGSANALALQSDARKSE 67
 Db 387 GYAPQGGGYAEPAAGRDYDYGQSGADPYQAPAGYISYGQGGYSGAGTSVTLQLDGSGR 446
 QY 68 TTITQSG---YNGADV-----GGADNSTIELTQNGFRNNATIDQ 105
 Db 447 TYQLREGSNITIGRGDAQFRLPDTGVSRHLEIRWDGQVALLADLNSTNGTTVNNAPVOE 506
 QY 106 WNAKNSDITVQGYDQVTVRVTHEMAHSAVMR 138
 Db 507 WQLADGDVY-----RLGHSEITVR 525

RESULT 10
 US-10-080-170-348
 ; Sequence 348, Application US/10080170
 ; Patent No. US20030129601A1
 ; GENERAL INFORMATION:
 ; APPLICANT: COLE, S.T.
 ; TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR
 ; TITLE OF INVENTION: IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR
 ; FILE REFERENCE: 03495,0218
 ; CURRENT APPLICATION NUMBER: US/10/080,170
 ; CURRENT FILING DATE: 2002-06-10
 ; PRIOR APPLICATION NUMBER: 60/270,123
 ; PRIOR FILING DATE: 2001-02-22
 ; NUMBER OF SEQ ID NOS: 652
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 348
 ; LENGTH: 527
 ; TYPE: PRT
 ; ORGANISM: Mycobacterium tuberculosis
 US-10-080-170-348

Query Match 11.2%; Score 86.5; DB 14; Length 527;
 Best Local Similarity 22.9%; Pred. No. 2.3;
 Matches 35; Conservative 13; Mismatches 56; Indels 49; Gaps 5;

QY 21 GVVPQWGG---GGNNHGGNSGSD-----STLSIYQGSANALALQSDARKSE 67
 Db 387 GYAPQGGGYAEPAAGRDYDYGQSGADPYQAPAGYISYGQGGYSGAGTSVTLQLDGSGR 446
 QY 68 TTITQSG---YNGADV-----GGADNSTIELTQNGFRNNATIDQ 105
 Db 447 TYQLREGSNITIGRGDAQFRLPDTGVSRHLEIRWDGQVALLADLNSTNGTTVNNAPVOE 506

QY 106 WNAKSDITVGGQIDQVTRVVTHEMAKSNVR 138
 Db 507 WOLADGDVI-----RLGHSSEIIVR 525

RESULT 11

US-10-156-761-8763
 ; Sequence 8763, Application US/10156761
 ; Publication No. US20030119018A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OMURA, SATOSHI
 ; APPLICANT: IKEDA, HARUO
 ; APPLICANT: ISHIKAWA, JUN
 ; APPLICANT: HORIKAWA, HIROSHI
 ; APPLICANT: SHIBA, TADAYOSHI
 ; APPLICANT: SAKAKI, YOSHIYUKI
 ; APPLICANT: HATTORI, MASHIIRA
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-262
 ; CURRENT APPLICATION NUMBER: US/10/156,761
 ; CURRENT FILING DATE: 2002-05-29
 ; PRIOR APPLICATION NUMBER: JP 2001-204089
 ; PRIOR FILING DATE: 2001-05-30
 ; PRIOR APPLICATION NUMBER: JP 2001-272697
 ; PRIOR FILING DATE: 2001-08-02
 ; NUMBER OF SEQ ID NOS: 15109
 ; SEQ ID NO 8763
 ; LENGTH: 482
 ; TYPE: PRT
 ; ORGANISM: Streptomyces avermitilis
 US-10-156-761-8763

Query Match

Best Local Similarity 11.1%; Score 86; DB 14; Length 482;
 Matches 36; Conservative 6; Mismatches 53; Indels 14; Gaps 4;

QY 8 AFAIVVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQVGSANALALQSDARKSE 67
 Db 256 AFAATSTFGS---GRAVFWGDSPIIDGTGSGNTLYDGNWDGATNALALNA----- 306
 QY 68 TTITGSGYNGADVGGADNSTIE--LTONGFRNNATTIDWNAKNSDIT 114
 Db 307 -TEMLVGASSGGDDGGGGGTCTAQLGNNHFGSGAT--TWTASSDVIT 352

RESULT 12

US-10-369-493-20619
 ; Sequence 20619, Application US/10369493
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 20619
 ; LENGTH: 486
 ; TYPE: PRT
 ; ORGANISM: Rhodopseudomonas palustris
 ; FEATURE:
 ; NAME/KEY: unsure
 ; LOCATION: (1) - (486)
 ; OTHER INFORMATION: unsure at all Xaa locations
 US-10-369-493-20619

Query Match 11.1%; Score 86; DB 15; Length 486;
 Best Local Similarity 24.4%; Pred. No. 2.3;
 Matches 29; Conservative 21; Mismatches 39; Indels 30; Gaps 5;

QY 11 AIVVSGS---ALAGVPPWGGGNNHNGGNSGPDSTLSIYQVGSANAL 57
 Db 86 SVVAGTDYRDYIAGVLPISLGGVKAAYVIDGSGPPAFPKDLASDPFSAPEFGADGFV 145
 QY 58 ALQSDARKSETTITGSGYNGADVGGADNSTIELTONGFRNNATTIDWNAKNSDITV 116
 Db 146 ITH-----TAAVGRPRGALISQ--NLLI-----AQSLLVAMRLTEADVNLG 187

RESULT 13

US-10-174-363-56
 ; Sequence 56, Application US/10174363
 ; Publication No. US20030077623A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Butler, Karlene H.
 ; APPLICANT: Cahoon, Rebecca E.
 ; APPLICANT: Harvell, Leslie T.
 ; APPLICANT: Rafaleki, Antoni J.
 ; APPLICANT: Sakai, Hajime
 ; TITLE OF INVENTION: Polynucleotides And Polypeptides Involved In Post-Transcriptional
 ; FILE REFERENCE: B1454 US NA
 ; CURRENT APPLICATION NUMBER: US/10/174,363
 ; CURRENT FILING DATE: 2002-06-17
 ; PRIOR APPLICATION NUMBER: 60/298,973
 ; PRIOR FILING DATE: 2001-06-18
 ; NUMBER OF SEQ ID NOS: 56
 ; SOFTWARE: Microsoft Office 97
 ; SEQ ID NO 56
 ; LENGTH: 1048
 ; TYPE: PRT
 ; ORGANISM: Arabidopsis thaliana
 US-10-174-363-56

Query Match 11.0%; Score 85.5; DB 14; Length 1048;
 Best Local Similarity 25.7%; Pred. No. 6.8;
 Matches 36; Conservative 17; Mismatches 60; Indels 27; Gaps 3;

QY 27 GGGGNNHNGGNSGSPSTLSI-----YQVGSANALALQSDARKSETTITGSGY 76
 Db 95 GGVGGRGSGSPSPQROSVPELHQTSTPTVAVSQPSTLSVSPQVEPTVLAQGF 153
 QY 77 NGADVGGADNSTIE-----LTONGFRNNATTIDWNAKNSDITVGOYDQ 120
 Db 154 EQLSVEGAPSAIQIPSSSKAFKPPMRPGKQSGKRCIVKANHFFAELPDIDLHRYDV 213
 QY 121 LVTRVVTHEMAKSNVRQV 140
 Db 214 TITPEVTSRGVNAVKOLV 233

RESULT 14

US-10-374-780A-2086
 ; Sequence 2086, Application US/10374780A
 ; Publication No. US2004001927A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Sherman, Bradley K.
 ; APPLICANT: Riechmann, Jose Luis
 ; APPLICANT: Jiang, Cai-Zhong
 ; APPLICANT: Heard, Jacqueline E.
 ; APPLICANT: Haake, Volker
 ; APPLICANT: Creelman, Robert A.
 ; APPLICANT: Ratcliffe, Oliver
 ; APPLICANT: Adam, Luc J.
 ; APPLICANT: Reuber, T. Lynne
 ; APPLICANT: Keddie, James
 ; APPLICANT: Brown, Pierre E.
 ; APPLICANT: Pilgrim, Marsha L.

APPLICANT: Dubell III, Arnold T
APPLICANT: Pineda, Omaira
APPLICANT: Yu, Guo-Liang
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES IN PLANTS
FILE REFERENCE: MB1-0047 CIP
CURRENT APPLICATION NUMBER: US/10/374,780A
CURRENT FILING DATE: 2003-02-25
PRIOR APPLICATION NUMBER: 09/837,944
PRIOR FILING DATE: 2001-04-18
PRIOR APPLICATION NUMBER: 60/310,847
PRIOR FILING DATE: 2001-08-09
PRIOR APPLICATION NUMBER: 09/334,455
PRIOR FILING DATE: 2001-08-22
PRIOR APPLICATION NUMBER: 60/336,049
PRIOR FILING DATE: 2001-11-19
PRIOR APPLICATION NUMBER: 60/338,692
PRIOR FILING DATE: 2001-12-11
PRIOR APPLICATION NUMBER: 10/171,468
PRIOR FILING DATE: 2002-06-14
PRIOR APPLICATION NUMBER: 10/225,066
PRIOR FILING DATE: 2002-08-09
PRIOR APPLICATION NUMBER: 10/225,067
PRIOR FILING DATE: 2002-08-09
PRIOR APPLICATION NUMBER: 10/225,068
PRIOR FILING DATE: 2002-08-09
NUMBER OF SEQ ID NOS: 2906
SOFTWARE: PatentIn version 3.2
SEQ ID NO 2086
LENGTH: 1048
TYPE: PRT
ORGANISM: Arabidopsis thaliana
FEATURE:
OTHER INFORMATION: G1152 Paralogous to G1146
US-10-374-780A-2086

Query Match 11.0%; Score 85.5; DB 15; Length 1048;
Best Local Similarity 25.7%; Pred. No. 6.8;
Matches 36; Conservative 17; Mismatches 60; Indels 27; Gaps 3;

QY 27 GGGGNNHGGGSSGPDSTLSI-----YQGSANALALQSDARKSETTITQSGYG 76
DB 95 GGGGGRGGSSGPPQSGVPELHQATSPYQAVSQPTLSEVSPTGVPEPTVLAQOF- 153
QY 77 NGADVGGAGDNSTIE-----LTQNGFRNNATTIDQNNKNSDITVGGYDQ 120
DB 154 EQLSYEGGAPSGAIGPISSSKAFKFPKPGGSGKRCIYKANHFFALPDKDLHRYDV 213
QY 121 LVTRVVTHEMAHASYMVRQV 140
DB 214 TTPEVTSRGVNRVAVMKQLV 233

RESULT 15
US-10-156-761-9343
Sequence 9343, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109

SEQ ID NO 9343
LENGTH: 438
TYPE: PRT
ORGANISM: Streptomyces avermitilis
US-10-156-761-9343

Query Match 11.0%; Score 85; DB 14; Length 438;
Best Local Similarity 25.3%; Pred. No. 2.6;
Matches 37; Conservative 22; Mismatches 51; Indels 36; Gaps 7;

QY 1 MKLIKVAFAFATVSGSALAGVVPQWGGGNNHGGSSGPDSTLSIYQGSANALAL-- 58
DB 1 MRSIRAAVGAATVMSUALAASAC-----GGSSSTGGSSNDSP-KTIT-YMASNGASIAVD 54
QY 59 ---LQSDARKSE-----TTITQSGYNGADVGGAGDNSTIELTQNG 96
DB 55 KKVLPPELDKFEHQGTGKVKLEVPWSDLLNRLTITTSGGGPDVANTIGNTSASIQATG 114
QY 97 FRNNATTIDQNNKNSDITVGGYDQLV 122
DB 115 ----ALLPWDANKFD-KIGKDRFV 134

Search completed: March 11, 2004, 19:18:35
Job time : 25.6 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:24:14 ; Search time 10.3 Seconds
(without alignments)
1410.186 Million cell updates/sec

Title: US-09-543-407-14

Perfect score: 775

Sequence: 1 MLKLKVAAPAAIVSGSALA.....HASVMVROYFGNNATANOY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	693	89.4	151	2	JC6039
2	693	89.4	151	2	AT0635
3	528	68.1	151	2	S70788
4	506.5	65.4	152	2	D90806
5	506.5	65.4	152	2	H85665
6	104.5	13.5	151	2	S70787
7	104.5	13.5	151	2	C90806
8	104.5	13.5	151	2	G85665
9	98.5	12.7	151	2	JC6040
10	98.5	12.5	552	2	D70604
11	96	12.4	770	2	TS1024
12	95	12.3	1748	2	S42136
13	93	12.0	2174	2	E95965
14	91	11.7	301	2	B84533
15	91	11.7	1028	2	AS6038
16	91	11.7	1213	2	SI6356
17	90.5	11.7	145	2	AD3143
18	90.5	11.7	145	2	H98144
19	90.5	11.7	151	2	AH0635
20	90	11.6	479	2	JN0891
21	89.5	11.5	256	2	T03171
22	89.5	11.5	262	2	S00275
23	89	11.5	479	2	C86266
24	88.5	11.4	573	2	A38307
25	87	11.2	590	2	E70946
26	87	11.2	646	1	SI9916
27	87	11.2	478	2	A70762
28	86.5	11.2	478	2	JN0892
29	86.5	11.2	527	2	B70700

30	86.5	11.2	645	2	F70825	probable ppe prote
31	86	11.1	447	2	G84687	probable disease r
32	86	11.1	599	2	B42049	leishmanolysin (EC
33	86	11.1	599	2	A44951	leishmanolysin (EC
34	85.5	11.0	575	2	S35327	protein kinase 899
35	85	11.0	347	2	B39112	merozoite 45K surf
36	85	11.0	967	2	S66852	hypothetical prote
37	84.5	10.9	440	2	AD1539	probable sugar ABC
38	84	10.8	407	2	T21956	hypothetical prote
39	84	10.8	590	1	A45621	leishmanolysin (EC
40	84	10.8	602	1	PL0221	leishmanolysin (EC
41	83.5	10.8	285	2	H84219	hypothetical prote
42	83.5	10.8	340	2	A83401	hypothetical prote
43	83.5	10.8	440	2	T15352	hypothetical prote
44	83.5	10.8	1067	2	S35423	protein kinase 899
45	83.5	10.8	3300	2	D70575	probable ppe prote

ALIGNMENTS

RESULT 1

JC6039 fimbria protein agfa precursor - Salmonella enteritidis

C/Species: Salmonella enteritidis

C/Date: 31-Dec-1996 #sequence revision 31-Dec-1996 #ext_change 08-Oct-1999

C/Accession: JC6039; PC6015; A44898

R/Collinson, S.K.; Cloutier, S.C.; Doran, J.L.; Baner, P.A.; Kay, W.W.

J. Bacteriol. 178, 662-667, 1996

A/Title: Salmonella enteritidis agfABC operon encoding thin, aggregative fimbriae.

A/Reference number: JC6039; PMID:96146512; PMID:8550497

A/Accession: JC6039

A/Molecule type: DNA

A/Residues: 1-151 <COL>

A/Cross-references: GB:U43280; NID:G184712; PID:AC43599.1; PID:G184714

A/Accession: PC6015

A/Molecule type: protein

A/Residues: 21-52 <CO2>

A/Experimental source: strain 27655-3b

A/Note: the authors translated the codon ACG for residue 44 as Ile

R/Collinson, S.K.; Emdy, L.; Muller, K.H.; Trust, T.J.; Kay, W.W.

J. Bacteriol. 173, 4773-4781, 1991

A/Title: Purification and characterization of thin, aggregative fimbriae from Salmonella

A/Reference number: A44898; PMID:91310586; PMID:1677357

A/Accession: A44898

A/Status: preliminary

A/Molecule type: protein

A/Residues: 21-33 <CO3>

A/Note: sequence extracted from NCBI backbone (NCBI:45936)

C/Genetics:

A/Genes: agfa

C/Function:

A/Description: major component of thin aggregative fimbriae

A/Note: fimbriae bind to fibronectin, plasminogen, tissue plasminogen activator

C/Keywords: fimbria

F/1-20/Domain: signal sequence #status predicted <SIG>

F/21-151/Product: fimbria protein agfa #status experimental <MAT>

Query Match

Best Local Similarity 89.4%; Score 693; DB 2; Length 151;

Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY	1	MLKLKVAAPAAIVSGSALAGVVPQWGGGNNHGGSSGPDSTLSIYOGSNAALALQ	60
DB	1	MLKLKVAAPAAIVSGSALAGVVPQWGGGNNHGGSSGPDSTLSIYOGSNAALALQ	60
QY	61	SDARKSETTITGSGYNGADVQAGADNSTIELTONGFRNNATIDQANASDITVCQYDQ	120
DB	61	SDARKSETTITGSGYNGADVQAGADNSTIELTONGFRNNATIDQANASDITVCQYDQ	120
QY	121	LVTRVYTHMAASVWVROYFGNNATANOY 151	
DB	121	LVTRVYTHMAASVWVROYFGNNATANOY 151	

Db 121 NNAAVNOQTASDSSVWVRQVFGNNATANQY 151

RESULT 2

AI0635

major curlin chain precursor [imported] - Salmonella enterica subsp. enterica serovar Typhimurium
 C/Species: Salmonella enterica subsp. enterica serovar Typhimurium
 A/Note: this species has also been called Salmonella typhimurium
 C/Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
 C/Accession: AI0635
 R/Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, R.; Croft, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Mout, S.; O'Garra, P.
 Nature 413, 848-852, 2001
 A/Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A/Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhimurium
 A/Reference number: AB0502; MUID:21534947; PMID:11677608
 A/Accession: AI0635
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-151 <PAR>
 A/Cross-References: GB:AL513382; PIDN:CAN08268.1; PID:G16502315; GSPDB:GN00176
 C/Genetics:
 A:Gene: STY1181

Query Match 89.4%; Score 693; DB 2; Length 151;
 Best Local Similarity 90.7%; Pred. No. 5e-52;
 Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 1 MLLKVAFAIIVSGSALAGVPPQMGGNNHGGSSGPDSTLSTIYGSANAAALQ 60
 DB 1 MLLKVAFAIIVSGSALAGVPPQMGGNNHGGSSGPDSTLSTIYGSANAAALQ 60
 QY 61 SDARSETTITQSGYNGADVQGGADNSTIELTQNGFRNATITDQWAKNSDITVGYDQ 120
 DB 61 SDARSETTITQSGYNGADVQGGADNSTIELTQNGFRNATITDQWAKNSDITVGYDQ 120
 QY 121 LVTRVYTHMAASVMTQVGFNNATANQY 151
 DB 121 NNAAVNOQTASDSSVWVRQVFGNNATANQY 151

RESULT 3

S70788

curlin protein csgA precursor - Escherichia coli (strain K-12)

N/Alternate names: csgA protein; major curlin protein

C/Species: Escherichia coli

C/Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 01-Mar-2002

C/Accession: S70788; G64846; S31202; S34560; S34559

R/Hammar, M.; Arngvist, A.; Bian, Z.; Olsen, A.; Normark, S.
 Mol. Microbiol. 18, 661-670, 1995

A/Title: Expression of two csg operons is required for production of fibronectin- and CsgA/Reference number: S70783; MUID:96414468; PMID:8817489

A/Accession: S70788

A/Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-151 <HAM>

A/Cross-References: EMBL:X90754; NID:G1147558; PIDN:CAA62282.1; PID:G1147564

A/Experimental source: strain K12, substrain W3110

A/Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995

R/Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co

A.; Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997

A/Title: The complete genome sequence of Escherichia coli K-12.

A/Reference number: A64720; MUID:97428617; PMID:9278503

A/Accession: G64846

A/Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-151 <BLAT>

A/Cross-References: GB:AE000205; GB:U00096; NID:G1787265; PIDN:AACT4126.1; PID:G1787279;
 R/Olsen, A.; Arngvist, A.; Hammar, M.; Sukupolvi, S.; Normark, S.
 Mol. Microbiol. 7, 523-536, 1993

A/Title: The Rpos sigma factor relieves H-NS-mediated transcriptional repression of csgA

A/Reference number: S31202; MUID:93211294; PMID:8459772

A/Accession: S31202

A/Molecule type: DNA

A/Residues: 1-6, 'V', '8-151 <OLS1>

A/Cross-References: EMBL:L04979

A/Accession: S34560

A/Molecule type: protein

A/Residues: 21-42; 44-50 <OLS2>

R/Olsen, A.N.; Arngvist, A.M.
 submitted to the EMBL Data Library, October 1992

A/Reference number: S34559

A/Accession: S34559

A/Molecule type: DNA

A/Residues: 1-133, 'KORDSGWLV' <OLS3>

A/Cross-References: EMBL:L04979; NID:G290424; PIDN:AAA23616.1; PID:G290425

A/Experimental source: strain K-12, substrain W3110

C/Genetics:

A:Gene: csgA

A:Map position: 23.15

C/Function:
 A/Description: major component of wild-type curli; interaction between CsgA and CsgB tr

A/Note: curli are thin, coiled fibers expressed on the surface of Escherichia coli that
 and H-kinogen; in the absence of CsgA, CsgB can self-assemble into polymers
 F;1-20/Domain: signal sequence #status predicted <SIG>
 F;21-151/Product: curlin #status experimental <MAT>

Query Match 68.1%; Score 528; DB 2; Length 151;
 Best Local Similarity 68.9%; Pred. No. 5.4e-38;
 Matches 104; Conservative 20; Mismatches 27; Indels 0; Gaps 0;

QY 1 MLLKVAFAIIVSGSALAGVPPQMGGNNHGGSSGPDSTLSTIYGSANAAALQ 60
 DB 1 MLLKVAFAIIVSGSALAGVPPQMGGNNHGGSSGPDSTLSTIYGSANAAALQ 60
 QY 61 SDARSETTITQSGYNGADVQGGADNSTIELTQNGFRNATITDQWAKNSDITVGYDQ 120
 DB 61 SDARSETTITQSGYNGADVQGGADNSTIELTQNGFRNATITDQWAKNSDITVGYDQ 120
 QY 121 LVTRVYTHMAASVMTQVGFNNATANQY 151
 DB 121 GNGAAVDQTASNSVNTQVGFNNATANQY 151

RESULT 4

D90806

curlin major subunit CsgA [imported] - Escherichia coli (strain O157:H7, substrain R1MD

C/Species: Escherichia coli

C/Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001

C/Accession: D90806

R/Hayashi, T.; Makino, K.; Onishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G
 dnaawa, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 Nara Res. 8, 11-22, 2001

A/Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen
 A/Reference number: A99629; MUID:21156231; PMID:11258796

A/Accession: D90806

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-152 <HAY>

A/Cross-References: GB:BA000007; PIDN:BA34843.1; PID:G13360880; GSPDB:GN00154
 A/Experimental source: strain O157:H7, substrain R1MD 0509952

C/Genetics:

A:Gene: ECs1420

Query Match 65.4%; Score 506.5; DB 2; Length 152;
 Best Local Similarity 67.1%; Pred. No. 3.6e-36;
 Matches 102; Conservative 21; Mismatches 28; Indels 1; Gaps 1;

QY 1 MLLKVAFAIIVSGSALAGVPPQMGGNNHGGSSGPDSTLSTIYGSANAAALQ 59
 DB 1 MLLKVAFAIIVSGSALAGVPPQMGGNNHGGSSGPDSTLSTIYGSANAAALQ 60
 QY 60 QSDARSETTITQSGYNGADVQGGADNSTIELTQNGFRNATITDQWAKNSDITVGYDQ 119

Db 61 QADARNSDLITTOHGGGNGADVGQSSDSDSLDTQGFNSATLDMNGKDSHMTVKQFG 120

QY 120 QLVTRVYTHMAHSAVMRVQVFGNNATANQY 151

Db 121 GGNAAVDDQTASNTVNTVQVFGNNATANHQY 152

RESULT 5

H85665

hypothetical protein csgA [imported] - Escherichia coli (strain O157:H7, substrain EDL933)

C/Species: Escherichia coli

C/Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #ext_change 14-Sep-2001

C/Accession: H85665

R/Perla, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca, Nature 409, 529-533, 2001

A/Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A/Reference number: AB5480; MUID:21074935; PMID:11206551

A/Accession: H85665

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-152 <STO>

A/Cross-references: GB:AE005174; NID:g12514574; PIDN:AA655788.1; GSPDB:GN00145; UMGF:216

A/Experimental source: strain O157:H7, substrain EDL933

C/Genetics:

A/Genes: csgA

Query Match 65.4%; Score 506.5; DB 2; Length 152;

Best Local Similarity 67.1%; Pred. No. 3.6e-36;

Matches 102; Conservative 21; Mismatches 28; Indels 1; Gaps 1;

QY 1 MLKLVAAPAAIVSGSALAGVPPQW-GGGGNNNGGNSGPDSTLSIYQGSANALAL 59

Db 1 MLKLKAAIAIAIFSSSALAGVPPQVGGGNNNGGNSGPNSELTNYQGGNSALAL 60

QY 60 QSDARSETTTTQSGYGNADVGQGDNSTIELTQNGFRNNATIDQNAKNSDITVQYD 119

Db 61 QADARNSDLITTOHGGGNGADVGQSSDSDSLDTQGFNSATLDMNGKDSHMTVKQFG 120

QY 120 QLVTRVYTHMAHSAVMRVQVFGNNATANQY 151

Db 121 GGNAAVDDQTASNTVNTVQVFGNNATANHQY 152

RESULT 6

570787

curlin nucleator protein csgB precursor - Escherichia coli (strain K-12)

N/Alternate names: csgB protein; curlin nucleation component; minor curlin protein

C/Species: Escherichia coli

C/Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #ext_change 01-Mar-2002

C/Accession: S70787; F64846

R/Hammer, M.; Aronqvist, A.; Bian, Z.; Olsen, A.; Normark, S. Mol. Microbiol. 18, 661-670, 1995

A/Title: Expression of two csg operons is required for production of fibronectin- and Cc

A/Reference number: S70783; MUID:96414468; PMID:8817489

A/Accession: S70787

A/Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-151 <HAM>

A/Cross-references: EMBL:X90754; NID:g1147558; PIDN:CAA62281.1; PID:g1147563

C/Species: Escherichia coli

C/Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #ext_change 14-Sep-2001

C/Accession: G85665

R/Perla, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca, Nature 409, 529-533, 2001

A/Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A/Reference number: AB5480; MUID:21074935; PMID:11206551

A/Accession: G85665

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-151 <STO>

A/Cross-references: GB:AE005174; NID:g12514574; PIDN:AA655787.1; GSPDB:GN00145; UMGF:21

A/Experimental source: strain K-12, substrain MG1655

C/Genetics:

A/Genes: csgB

A/Map position: 23.15

C/Function:

A/Description: minor component of wild-type curli; interaction between CsgA and CsgB tr

A/Note: curli are thin, coiled fibers expressed on the surface of Escherichia coli that

and H-kinogen; in the absence of CsgA, CsgB can self-assemble into polymers

F/1-21/Domain: signal sequence #status predicted <SIG>

F/22-151/Product: minor curlin chain #status predicted <MAT>

Query Match 13.5%; Score 104.5; DB 2; Length 151;

Best Local Similarity 29.2%; Pred. No. 0.054;

Matches 35; Conservative 19; Mismatches 59; Indels 7; Gaps 4;

QY 9 FAIVVSGSAL--AGVPQWGGGNNNGGNSGPDSTLSIY-QYGSANALALQSDARK 65

Db 33 FAVNELSKSFPQAAIIT---GQAGTNSAQLQGGSKLAAVAAQESSNRA-KIDQTCGY 88

QY 66 SETTTQSGYGNADVGQGDNSTIELTQNGFRNNATIDQNAKNSDITVQYDQLVTRV 125

Db 89 NLAVIDQASANDASISQAGYNTAMIIQKSGNKNKITQYGTQKTAIVVQROSQMAIRV 148

RESULT 7

C90806

minor curlin subunit precursor CsgB [imported] - Escherichia coli (strain O157:H7, sub-

C/Species: Escherichia coli

C/Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #ext_change 18-Jul-2001

C/Accession: C90806

R/Hayashi, T.; Makino, K.; Onishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G

gatawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H. DNA Res. 8, 11-22, 2001

A/Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen

A/Reference number: A99629; MUID:21156231; PMID:11258796

A/Accession: C90806

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-151 <HAY>

A/Cross-references: GB:BA000007; PIDN:BA34842.1; PID:g13360879; GSPDB:GN00154

A/Experimental source: strain O157:H7, substrain RIMD 0509952

C/Genetics:

A/Genes: EcsA19

Query Match 13.5%; Score 104.5; DB 2; Length 151;

Best Local Similarity 29.2%; Pred. No. 0.054;

Matches 35; Conservative 19; Mismatches 59; Indels 7; Gaps 4;

QY 9 FAIVVSGSAL--AGVPQWGGGNNNGGNSGPDSTLSIY-QYGSANALALQSDARK 65

Db 33 FAVNELSKSFPQAAIIT---GQAGTNSAQLQGGSKLAAVAAQESSNRA-KIDQTCGY 88

QY 66 SETTTQSGYGNADVGQGDNSTIELTQNGFRNNATIDQNAKNSDITVQYDQLVTRV 125

Db 89 NLAVIDQASANDASISQAGYNTAMIIQKSGNKNKITQYGTQKTAIVVQROSQMAIRV 148

RESULT 8

G85665

curlin minor chain precursor, CsgA homolog [imported] - Escherichia coli (strain O157:H

C/Species: Escherichia coli

C/Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #ext_change 14-Sep-2001

C/Accession: G85665

R/Perla, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca, Nature 409, 529-533, 2001

A/Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A/Reference number: AB5480; MUID:21074935; PMID:11206551

A/Accession: G85665

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-151 <STO>

A/Cross-references: GB:AE005174; NID:g12514574; PIDN:AA655787.1; GSPDB:GN00145; UMGF:21

A;Accession: S421

A.Molecule type: DNA
A.Residues: 1-1748 <TAV>
A.Cross-references: EMBL:L03710; NID:g161751; PID:g161752
R.Taylor, F.M.; Martindale, D.W.
Nucleic Acids Res. 21, 4610-4614, 1993
A.Title: Retroviral-type zinc fingers and glycine-rich repeats in a protein encoded by c

A.Accession: S42135
A.Molecule type: DNA
A.Residues: 1164-1174,1179-1199,1233-1252,1285-1293,1297-1309,1316-1326,1331-1341,1343-1

A.Cross-references: EMBL:L03710
R.Martindale, D.W.; Taylor, F.M.
Nucleic Acids Res. 16, 2189-2201, 1988
A>Title: Multiple introns in a conjugation-specific gene from Tetrahymena thermophila.

A.Reference number: S03650; MUID:88189811; PMID:3357771
A.Molecule type: DNA
A.Residues: 236-250, 'I', 252-255, 'N', 257-773 <MAR>

A.Cross-references: EMBL:X06462
C.Genetics:
A.Gene: cnjB
A.Genetic code: SGCS
A.Introns: 85/3; 136/1; 157/3; 201/2; 290/2; 327/3; 499/1; 573/2; 607/3; 708/3; 777/3; 8

C.Keywords: zinc finger
P.1154-1450/Region: glycine-rich
F.1451-1464/Region: zinc finger CCHC motif
F.1478-1491/Region: zinc finger CCHC motif
F.1501-1514/Region: zinc finger CCHC motif
F.1530-1543/Region: zinc finger CCHC motif
F.1555-1568/Region: zinc finger CCHC motif
F.1579-1592/Region: zinc finger CCHC motif
F.1602-1615/Region: zinc finger CCHC motif
F.1626-1748/Region: glycine-rich

Query Match 12.3%; Score 95; DB 2; Length 1748;
Best Local Similarity 32.0%; Pred. No. 5.6;
Matches 31; Conservative 11; Mismatches 29; Indels 26; Gaps 4;

OY 25 QWGGGNNHNG--GNSSGPSTLSIVQGSANALALQSDARKSETIT---OSGYCN 77
|::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1640 QPGCGNSNGCGMGWGSSGDWN-----CQSNVSESTTTSSCGMSSGSGN 1685

OY 78 GADVGGADNSTELTONGFRNNATIDQMNASDIT 114
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1686 QTGGCGWSNDN-----QQQNENGTGGCGMSSSNGQT 1717

RESULT 13
E95965
hypothetical glycine-rich protein [imported] - Sinorhizobium meliloti (strain 1021) mega
C.Species: Sinorhizobium meliloti
C.Date: 24-Aug-2001 #sequence_reviseion 24-Aug-2001 #text_change 30-Sep-2001
CAccession: E95965
R.Filman, T.M.; Weidner, S.; Wong, K.; Buhrmeester, J.; Chain, P.; Vorholter, F.U.; Herman
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A>Title: The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing endo
A.Reference number: A95842; MUID:21396508; PMID:11481431
A.Accession: E95965
A>Status: preliminary
A.Molecule type: DNA
A.Residues: 1-2174 <KUR>
A.Cross-references: GB:AL591985; PIDN:CAC49389.1; PID:g15140875; GSDB:GN00167
A.Experimental source: strain 1021, megaplasmid PSymb
R.Galibert, F.; Filman, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A.Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelauré,
heault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A>Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A.Reference number: A96039; MUID:21368234; PMID:11474104
A.Contents: annotation
C.Genetics:

A:Gene: SMD215158
A:Genome: plasmid

Query Match 12.0%; Score 93; DB 2; Length 2174;
Best Local Similarity 24.5%; Pred. No. 11;
Matches 36; Conservative 15; Mismatches 32; Indels 64; Gaps 7;

QY 11 ALVVSGLAGVAVQ--WGGGNNHGGSSGSPSTLSIYQGSANMALAQSDARKSET 68
DB 693 AATAGACAGAVIILAQSIQGGGNN--GGMATGGAGAGSGFQIGGGGGG----- 737
QY 69 TITQSGYNGADVQ--QGANSTIELTONGFRNNATTIDQWNAKNSDITQGYDLVTRVVT 127
DB 738 ---GGYNTANTAVGFKG-----LTLTTGG----- 757
QY 128 HEMAHASVYVQV---GFGNNATANQY 151
DB 758 ---SHAAGIVAQSVGGGGGTGTASSY 781

RESULT 14
B84533
hypothetical protein At2g15770 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: B84533
R:Lin, X.; Kaul, S.; Roundley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffett, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: B84533
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 15301 <STO>
A:Cross-references: GB:AE002093; NID:g5306254; PIDN:AAD41987.1; GSPDB:GN00139
A:Genetic8:
A:Gene: At2g15770
A:Map position: 2

Query Match 11.7%; Score 91; DB 2; Length 301;
Best Local Similarity 26.4%; Pred. No. 1.7;
Matches 33; Conservative 19; Mismatches 51; Indels 22; Gaps 5;

QY 15 SGSALA-----GVVPQMGCGGNH--NGGNSGSPDSTLSIYQGSANMALAQSDARKSET 68
DB 69 SSGSLGNSNNNGSGMGWGTGNSRSGSGSSTNPGRSRSMW-----SLKSGMSWKG 121
QY 69 TITQSGYNGADVQGANST---IELTONGFRNNATTIDQWNAKNSDITVGYD 119
DB 122 SINDNSNSGSGDSSGLDRETPKKIIVGSGDWKGLDYDQMAKNAFVYNDVLVRYD 181
QY 120 QLVTR 124
DB 182 KSAKR 186

RESULT 15
A56038
DNA-binding protein ovo - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 01-Dec-1995 #sequence_revision 01-Dec-1995 #text_change 21-Jul-2000
C:Accession: A56038
R:Garfinkel, M.D.; Wang, J.; Liang, Y.; Mahowald, A.P.
Mol. Cell. Biol. 14, 6809-6818, 1994
A:Title: Multiple products from the shavenbaby-ovo gene region of Drosophila melanogaster.
A:Reference number: A56038; MUID:95021209; PMID:7935398
A:Accession: A56038
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1028 <GAR>
A:Cross-references: GB:U11363; NID:g520526; PIDN:AAB60216.1; PID:g520527

C:Genetics:
A:Gene: ovo
A:Cross-references: FlyBase:FBgn0003028

Query Match 11.7%; Score 91; DB 2; Length 1028;
Best Local Similarity 31.2%; Pred. No. 6.7;
Matches 34; Conservative 8; Mismatches 41; Indels 26; Gaps 4;

Qy 3 LKVAFAAIVVSGSALAGVFPQWGGGNNHGGNSGPPDSTLSTIYQGSANAALALQSD 62
Db 59 LONAAAAYIMSAGS-----GGGCTGGGGGASGPGGGSANSGGGGGG----- 104
Qy 63 ARKSETTITQSGYGNGADVGGAGDNSTIELTONGFRNNATIDQWNAKNS 111
Db 105 -----CGNGYINGGIVG-GPNNS---LDGNNLINFASVSNSNESNS 141

Search completed: March 11, 2004, 18:42:07
Job time : 11.3 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Comugen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:14:48 ; Search time 6.3 Seconds

(without alignments)
1248.031 Million cell updates/sec

Title: US-09-543-407-14

Perfect score: 775
Sequence: 1 MRLKLVAFAPAIIVSGSALA.....HASVMVRYQFGNNATANY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	693	89.4	151	1 CSGB_SALTY	P55225 salmonella
2	528	68.1	151	1 CSGB_ECOLI	P28307 escherichia
3	506.5	65.4	152	1 CSGB_ECO57	Q93U24 escherichia
4	104.5	13.5	151	1 CSGB_ECOLI	P39828 escherichia
5	98.5	12.7	151	1 CSGB_SALTY	P55226 salmonella
6	92	11.9	1327	1 TNKI_HUMAN	O95271 homo sapien
7	91	11.7	1028	1 OVO_DROME	P51521 drosophila
8	90.5	11.7	151	1 CSGB_SALTY	Q82783 salmonella
9	90	11.6	1093	1 PER_DROMI	Q03297 drosophila
10	89.5	11.5	262	1 VG38_BPT2	P07875 bacterioph
11	88.5	11.4	479	1 PRIC_BRMCH	P16317 erwina chr
12	87	11.2	646	1 GP63_LEIME	P43150 leishmania
13	87	11.2	678	1 YP48_MYCTU	Q10778 mycobacteri
14	86	11.1	599	1 GP63_LEICH	P15706 leishmania
15	85.5	11.0	1048	1 AGO1_ARATH	O04379 arabidopsis
16	85	11.0	347	1 MSAG_PLAF2	O03646 plasmidom
17	84	10.8	590	1 GP63_LEIDO	P23223 leishmania
18	84	10.8	602	1 GP63_LEIMA	P08148 leishmania
19	83.5	10.8	440	1 PD46_CABEL	Q11067 caenorhabd
20	83.5	10.8	1067	1 SGG_DROME	P18431 drosophila
21	83	10.7	362	1 P35_MYCPE	Q50367 mycoplasma
22	83	10.7	1656	1 OMPB_RICJA	O06653 r outer mem
23	82	10.6	1080	1 HDC_DROME	O92268 drosophila
24	81.5	10.5	392	1 HME1_HUMAN	O59525 homo sapien
25	81	10.5	165	1 GRP1_ORYSA	P25074 oryza sativ
26	81	10.5	1612	1 RRPD_PMVU	P89657 pepper mild
27	80.5	10.4	1612	1 RRPD_PMVU	P29098 pepper mild
28	80.5	10.4	481	1 PRPB_BRMCH	P16316 erwina chr
29	80.5	10.4	641	1 IMD_ARIGO	O44052 arthropacte
30	80	10.3	385	1 PER_DROME	P91686 drosophila
31	80	10.3	491	1 YK98_MYCTU	Q10707 mycobacteri
32	80	10.3	548	1 CEAK_ECOLI	Q47502 escherichia
33	79.5	10.3	204	1 HEVE_HEVER	P02877 hevea brasl

34	79	10.2	396	1 PER_DROMI	P91698 drosophila
35	79	10.2	933	1 NPA3_HUMAN	Q81410 homo sapien
36	78.5	10.1	1567	1 ICEN_XANCT	P18127 xanthomonas
37	78	10.1	331	1 OMB2_NEIMB	P30688 neisseria m
38	78	10.1	576	1 DEAF_DROME	Q24180 drosophila
39	77.5	10.0	1140	1 YMG6_YEAST	O04893 saccharomyc
40	77	9.9	401	1 YK03_CABEL	P34291 caenorhabd
41	77	9.9	783	1 YK02_CABEL	P34308 caenorhabd
42	77	9.9	796	1 MBN_DROME	P52302 drosophila
43	77	9.9	1690	1 KFL1_HUMAN	O12756 homo sapien
44	76.5	9.9	329	1 SUP1_ECOLI	P06997 escherichia
45	76.5	9.9	1115	1 PHYE_PLANT	P55004 pharbitis n

ALIGNMENTS

RESULT 1	CSGB_SALTY	STANDARD;	PRT;	151 AA.
AC	P55225;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Major curilin subunit precursor (Fimbrin SEF17).			
GN	CSGB OR AGFA OR STM1144 OR STY1181 OR T1776.			
OS	Salmonella typhimurium,			
OS	Salmonella enteritidis.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OC	Enterobacteriaceae; Salmonella.			
NCBI_Taxid=602, 601, 592;				
RP	SEQUENCE FROM N.A.			
RC	SPECIES=S.typhimurium; STRAIN=SR-11;			
RX	MEDLINE=98117058; PubMed=9457880;			
RA	Romling U., Bian Z., Hammar M., Sierra W.D., Normark S.;			
RT	"Curli fibers are highly conserved between Salmonella typhimurium and			
RT	Escherichia coli with respect to operon structure and regulation.";			
RL	J. Bacteriol. 180:722-731(1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;			
RA	McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,			
RA	Courtney L., Bernwolk S., Ali J., Dante M., Du F., Hou S., Layman D.,			
RA	Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,			
RA	Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,			
RT	Waterston R., Wilson R.K.;			
RT	"Complete genome sequence of Salmonella enterica serovar Typhimurium			
RT	LT2.";			
RL	Nature 413:852-856(2001).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RX	SPECIES=S.typhimurium; STRAIN=CT18;			
RX	MEDLINE=21534947; PubMed=11677608;			
RA	Parhill T., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,			
RA	Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebatina M.,			
RA	Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,			
RA	Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,			
RA	Fellwell T., Hamlin N., Haque A., Hien T.T., Holtroyd S., Jagels K.,			
RA	Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,			
RA	Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,			
RA	Whitehead S., Barrett B.G.;			
RT	"Complete genome sequence of a multiple drug resistant Salmonella			
RT	enterica serovar Typhimurium CT18.";			
RL	Nature 413:848-852(2001).			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RX	SPECIES=S.typhimurium; STRAIN=Ty2 / ATCC 700931;			
RX	MEDLINE=22531367; PubMed=12644504;			
RA	Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,			
RA	Burford V., Kodoyanni V., Schwartz D.C., Blattner F.R.;			

RT "Comparative genomics of *Salmonella enterica* serovar Typhimurium strains Ty2
 RT and CT18.";
 RL J. Bacteriol. 185:2330-2337(2003).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S.enteritidis; STRAIN=27655-3B;
 RX MEDLINE=96146512; PubMed=8550497;
 RA Collinson S.K., Clouthier S.C., Doran J.L., Baner P.A., Kay W.W.,
 RT "Salmonella enteritidis agfBAC operon encoding thin, aggregative
 RT fimbriae.";
 RL J. Bacteriol. 178:662-667(1996).
 RN [6]
 RP SEQUENCE OF 21-151 FROM N.A.
 RC SPECIES=S.enteritidis; STRAIN=27655-3B;
 RX MEDLINE=94013373; PubMed=8104955;
 RA Doran J.L., Collinson S.K., Burtian J., Santos G., Todd E.C.D.,
 RT Munro C.K., Kay C.M., Baner P.A., Peterkin P.I., Kay W.W.,
 RT "DNA-based diagnostic tests for *Salmonella* species targeting agfA,
 RT the structural gene for thin, aggregative fimbriae.";
 RL J. Clin. Microbiol. 31:2263-2273(1993).
 RN [7]
 RP SEQUENCE OF 21-33.
 RC SPECIES=S.enteritidis; STRAIN=27655-3B;
 RX MEDLINE=91310586; PubMed=1677357;
 RA Collinson S.K., Emsley L., Mueller K.-M., Trust T.J., Kay W.W.,
 RT "Purification and characterization of thin, aggregative fimbriae from
 RT *Salmonella enteritidis*.";
 RL J. Bacteriol. 173:4773-4781(1991).
 CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 CC FIBRONECTIN.
 CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AJ0023101; CAA05317.1; -
 DR EMBL: AE008749; AAL20074.1; -
 DR EMBL: AE016840; CAD08268.1; -
 DR EMBL: AE016840; AAO69399.1; -
 DR EMBL: U43280; AAC43599.1; -
 DR PIR: JC6039; JC6039.
 DR StyGene: SGI0608; CSGA.
 KW Fimbria; Signal; Complete proteome.
 FT CHAIN 1 20
 FT SIGNAL 1 20
 FT CHAIN 1 151
 FT CONFLICT 134 151
 FT REF. 6)
 SQ SEQUENCE 151 AA; 15305 MW; B7DAC0D16B621359 CRC64;
 Query Match 89.4%; Score 693; DB 1; Length 151;
 Best Local Similarity 90.7%; Pred. No. 4.9e-53;
 Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPWVGNGGNGGSGPDSPTISYVGSANALALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPWVGNGGNGGSGPDSPTISYVGSANALALQ 60
 QY 61 SDAKSTTTTOSGYGADVGAGADNSTELTQNGFRNATIDWNKAKSDITVGOYDQ 120
 DB 61 SDAKSTTTTOSGYGADVGAGADNSTELTQNGFRNATIDWNKAKSDITVGOYDQ 120
 QY 121 LTVRVVTHEMAHASVWVROYGFGNNATANY 151
 DB 121 LTVRVVTHEMAHASVWVROYGFGNNATANY 151
 QY 121 NNAALVIVQTAZSSVWVROYGFGNNATANY 151
 DB 121 NNAALVIVQTAZSSVWVROYGFGNNATANY 151

RESULT 2
 ID CSGA_ECOLI STANDARD; PRT; 151 AA.
 AC P28307;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Major curlin subunit precursor.
 GN CSGA OR B1042.
 OS *Escherichia coli*.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; *Escherichia*.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / W3110;
 RX MEDLINE=93211294; PubMed=8459772;
 RA Olsen A., Argyris A.,
 RT "The Rpos sigma factor relieves H-NS-mediated transcriptional
 RT repression of *csaA*, the subunit gene of fibronectin-binding curli in
 RT *Escherichia coli*.";
 RL Mol. Microbiol. 7:523-536(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / WC4100;
 RX MEDLINE=96414468; PubMed=8817489;
 RA Hammer M., Argyris A., Bian Z., Olsen A., Normark S.,
 RT "Expression of two *csa* operons is required for production of
 RT fibronectin- and Congo red-binding curli polymers in *Escherichia coli*
 RT K-12.";
 RL Mol. Microbiol. 18:661-670(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna C.A., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.,
 RT "The complete genome sequence of *Escherichia coli* K-12.";
 RL Science 277:1234-1238(1997).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=9706102; PubMed=8905232;
 RA Oshima T., Alpha H., Baba T., Fujita K., Hayashi K., Honjo A.,
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
 RA Sasaki G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horikuchi T.,
 RT "A 718-kb DNA sequence of the *Escherichia coli* K-12 genome
 RT corresponding to the 12.7-28.0 min region on the linkage map.";
 RL DNA Res. 3:137-155(1996).
 RN [5]
 RP SEQUENCE OF 21-40.
 RC STRAIN=K12 / YMBL;
 RX MEDLINE=9302873; PubMed=1357528;
 RA Argyris A., Olsen A., Pfeiffer J., Russell D.G., Normark S.,
 RT "The Csl protein activates cryptic genes for curli formation and
 RT fibronectin binding in *Escherichia coli* HB101.";
 RL Mol. Microbiol. 6:2443-2452(1992).
 RN [6]
 RP SEQUENCE OF 21-31.
 RC MEDLINE=91310586; PubMed=1677357;
 RA Collinson S.K., Emsley L., Trust T.J., Kay W.W.,
 RT "Purification and characterization of thin, aggregative fimbriae from
 RT *Salmonella enteritidis*.";
 RL J. Bacteriol. 173:4773-4781(1991).
 CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 CC FIBRONECTIN.

```

CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L04979; AAA3616.1; -
DR EMBL; X80754; CAA62282.1; -
DR EMBL; AE00205; AAC74126.1; -
DR EMBL; D90741; BAA35832.1; -
DR EMBL; D90742; BAA35840.1; -
DR PIR; S70788; S70788.
DR Ecocore; EC11489; csgr.
KM Fimbrtia; Signal; Complete proteome.
FT SIGNAL 1 20
FT CHAIN 21 151
FT CONFID 7 7
SQ SEQUENCE 151 AA; 15049 MW; C003470D208D395F CRC64;

Query Match 68.1%; Score 528; DB 1; Length 151;
Best Local Similarity 68.9%; Pred. No. 7, 9e-39;
Matches 104; Conservative 20; Mismatches 27; Indels 0; Gaps 0;

QY 1 MLTLVAAFAAIVSGSALAGVPPWPGGGGNNHGGNSGSPSTLSIYQGSANALALQ 60
DB 1 MLTLVAAFAAIVFSSSALAGVPPQYGGGNGHGGNSGNSPSELNITVYQGGNSALALQ 60
QY 61 SDARKSETTITQSGYNGADVGGADNSTIELTONGFRNNATTIDQNNAKNSDITVGYDQ 120
DB 61 TDARNSDLITTHGGNGADVGGSDSSIDLTRGFNGSATILDQNNKNSSEITVQPGG 120

QY 121 LVTRVVTHEMAHSAVMVRQVGFNNATNAY 151
DB 121 GNGAAVDTASNSVTVQVFGNNATNAYQ 151

RESULT 3
CSGB_ECO57 STANDARD; PRT; 152 AA.
AC 093U24;
ID CSGA_ECO57
AD 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Major curlin subunit precursor.
CSG OR Z1676 OR ECS1420.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_Taxid=83334;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / ATCC 43895;
RX MEDLINE=21218556; PubMed=11319125;
RA Uhlisch G.A., Keen J.E., Elder R.O.;
RT "Mutations in the csgrD promoter associated with variations in curli
RT expression in certain strains of Escherichia coli O157:H7.";
RL Appl. Environ. Microbiol. 67:2367-2370(2001).
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).

```

```

RN (3)
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Onoebudo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,
RA Kihara S., Shiba T., Hatori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
RN (4)
RP FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
RP COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
RP TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
RP FIBRONECTIN.
CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF275733; AAK53212.1; -
DR EMBL; AE005315; AAG55788.1; -
DR EMBL; AP002554; BAB34843.1; -
DR PIR; D90806; D90806.
DR PIR; H85655; H85655.
KM Fimbrtia; Signal; Complete proteome.
FT SIGNAL 1 20
FT CHAIN 21 152
SQ SEQUENCE 152 AA; 15099 MW; EE2D2094DDE91243 CRC64;

Query Match 65.4%; Score 506.5; DB 1; Length 152;
Best Local Similarity 67.1%; Pred. No. 5, 6e-37;
Matches 102; Conservative 21; Mismatches 28; Indels 1; Gaps 1;

QY 1 MLTLVAAFAAIVSGSALAGVPPWPGGGGNNHGGNSGSPSTLSIYQGSANALALQ 59
DB 1 MLTLVAAFAAIVFSSSALAGVPPQYGGGNGHGGNSGNSPSELNITVYQGGNSALALQ 60
QY 60 OSDARKSETTITQSGYNGADVGGADNSTIELTONGFRNNATTIDQNNAKNSDITVGYDQ 119
DB 61 QADARNSDLITTHGGNGADVGGSDSSIDLTRGFNGSATILDQNNKNSSEITVQPGG 120

QY 120 QLVTRVVTHEMAHSAVMVRQVGFNNATNAY 151
DB 121 GNGAAVDTASNSVTVQVFGNNATNAYQ 152

RESULT 4
CSGB_ECOLI STANDARD; PRT; 151 AA.
AC P39828;
ID CSGA_ECOLI
AD 01-FEB-1995 (Rel. 31, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Minor curlin subunit precursor.
CSG OR B1041 OR Z1675 OR ECS1419.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_Taxid=562, 83334;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MC4100;
RX MEDLINE=96414468; PubMed=8817489;
RA Hammar M., Arrqvist A., Bian Z., Olsen A., Normark S.;
RT "Expression of two csg operons is required for production of
RT fibronectin- and csg red-binding curli polymers in Escherichia coli

```

RT K-12." ;
 RL Mol. Microbiol. 18:661-670 (1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blatter F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y. ;
 RT "The complete genome sequence of *Escherichia coli* K-12." ;
 RL Science 277:1453-1474 (1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97061202; PubMed=8905232;
 RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
 RA Sampaio G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horiiuchi T. ;
 RT "A 718-kb DNA sequence of the *Escherichia coli* K-12 genome
 corresponding to the 12.7-28.0 min region on the linkage map." ;
 RL DNA Res. 3:137-155 (1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller E.,
 RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamocostas K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blatter F.R. ;
 RT "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7." ;
 RL Nature 409:529-533 (2001).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,
 RA Kihara S., Shiba T., Hattori M., Shinagawa H. ;
 RT "Complete genome sequence of enterohaemorrhagic *Escherichia coli*
 O157:H7 and genomic comparison with a laboratory strain K-12." ;
 RL DNA Res. 8:111-22 (2001).
 RN [6]
 RP SEQUENCE OF 1-21 FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=95157246; PubMed=7854117;
 RA Arngvist A., Olsen A., Normark S. ;
 RT "Sigma S-dependent growth-phase induction of the *csgBA* promoter in
Escherichia coli can be achieved in vivo by sigma 70 in the absence
 of the nucleoid-associated protein H-NS." ;
 RL Mol. Microbiol. 13:1021-1032 (1994).
 RN [7]
 RP COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
 CURLIN MONOMERS.
 RN [8]
 RP SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.

DR EMBL, AE000205; AAC74125.1; -
 DR EMBL, D90741; BAA35831.1; -
 DR EMBL, AE005315; AAG55787.1; -
 DR EMBL, AP002554; BAB34842.1; -
 DR PIR, C90806; C90806.
 DR PIR, G85665; G85665.
 DR PIR, S70787; S70787.
 DR Ecocyc, EG12621; csgB.
 KW Fimbrin; Signal; Complete proteome.
 FT SIGNAL 1 21
 FT CHAIN 22 151
 SQ SEQUENCE 151 AA; 15882 MW; B18D268B964014B8 CRC64;
 Query Match 13.5%; Score 104.5; DB 1; Length 151;
 Best Local Similarity 29.2%; Pred. No. 0.023;
 Matches 35; Conservative 19; Mismatches 59; Indels 7; Gaps 4;
 QY 9 FAATVSGSAL-AGVVPQWGGGNGHNGSSGPDSTLSIY-QYGSANALALQSPARK 65
 DB 33 FAVNELSKSPNQALII---GQAGTNNNAQLRQGGSKLAVVAQEGSGNRA-KIDQTGDY 88
 QY 66 SETTIGSGYGVGADVGGADNSTELTQNGFRNNATIDQWNAKNSDITGVQYDQVTRV 125
 DB 89 NNAVTDQAGSANDASISQAGYNTAMIIQKSGKNVANTVGTCTKTAIVQROSGQNAIRV 148
 RESULT 5
 CSGB_SALTY STANDARD; PRT; 151 AA.
 ID CSGB_SALTY P55256;
 AC P55256;
 DT 01-OCT-1996 (rel. 34, Created)
 DT 01-OCT-1996 (rel. 34, Last sequence update)
 DT 28-FEB-2003 (rel. 41, Last annotation update)
 DE Minor curli subunit precursor (Fimbrin SEF17 minor subunit).
 GN CSGB OR AGFB OR STM1143.
 OS Salmonella typhimurium; and
 OS Salmonella enteritidis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 OX NCBI_Taxid=602, 592;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S.typhimurium; STRAIN=SR-11;
 RX MEDLINE=98117058; PubMed=9457880;
 RA Romling U., Bian Z., Hammar M., Sierraalta W.D., Normark S. ;
 RT "Curli fibers are highly conserved between *Salmonella typhimurium* and
Escherichia coli with respect to operon structure and regulation." ;
 RL J. Bacteriol. 180:722-731 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
 RX MEDLINE=21534948; PubMed=11677609;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
 RA Waterston K., Wilson R.K. ;
 RT "Complete genome sequence of *Salmonella enterica* serovar *Typhimurium*
 LT2." ;
 RL Nature 413:852-856 (2001).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S.enteritidis; STRAIN=27655-3B;
 RX MEDLINE=96146512; PubMed=8550497;
 RA Collinson S.K., Clouthier S.C., Doran J.L., Baner P.A., Kay W.W. ;
 RT "Salmonella enteritidis *agfBAC* operon encoding thin, aggregative
 fimbriae." ;
 RL J. Bacteriol. 178:662-667 (1996).
 RN [4]
 RP FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
 CURLIN MONOMERS.

CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; A1002301; CA05316.1; -
 DR EMBL; AE008749; AAL20073.1; -
 DR EMBL; U43280; AAC33598.1; -
 DR PIR; JC6040; JC6040.
 DR StyGene; SG10609; CSGB.
 DR Fimbrla; Signal; Complete proteome.
 FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 1 21 MINOR CURLIN SUBUNIT.
 FT SEQUENCE 151 AA; 16182 MW; C0FC5430E6DD361D CRC64;
 SQ
 Query Match 12.7%; Score 98.5; DB 1; Length 151;
 Best Local Similarity 26.2%; Pred. No. 0.074;
 Matches 34; Conservative 17; Mismatches 42; Indels 37; Gaps 4;
 QY 10 AAIWVSGALAGVPGWGGGNNHGGSSGPDSTLSTIYQGSANALALQSDARKSETT 69
 DB 58 AAVRQGSXKLIVISQ-EGNNRAKYDQAGNNFAYIEQTGNAN-----DAS 103
 QY 70 ITQSGYGNADVQGGADNSTIELTONGFNNAITIDPMANKNSITTYGQDQVTRVYTHE 129
 DB 104 IQSQAAGNSA-----AIQKSGSKANITTYGTOK-----TAVVQK 140
 QY 130 MAHAYVVRQ 139
 DB 141 QSHMARVQ 150
 RESULT 6
 TNK1_HUMAN STANDARD; PRT; 1327 AA.
 ID TNK1_HUMAN
 AC O95271; O95272.
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Tankyrase 1 (EC 2.4.2.30) (TNK1) (Tankyrase I) (TNKS-1) (TRF1-
 DE interacting ankyrin-related ADP-ribose polymerase).
 GN TNKS OR TNK1 OR TINI OR TINP1 OR PARP.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OC NCBI_TaxID=9606;
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
 RC TISSUE=Testis;
 RA MEDLINE=99040105; PubMed=9822378;
 RA Smith S., Giriat L., Schmitt A., de Lange T.;
 RT "Tankyrase, a poly(ADP-ribose) polymerase at human telomeres.";
 RL Science 282:1484-1487(1998).
 RN [2]
 RP SUBCELLULAR LOCATION.
 RX MEDLINE=99454782; PubMed=10523501;
 RA Smith S., de Lange T.;
 RT "Cell cycle dependent localization of the telomeric PARP, tankyrase,
 RT to nuclear pore complexes and centrosomes.";
 RL J. Cell Sci. 112:3649-3656(1999).
 RN [3]
 RP FUNCTION, AND PHOSPHORYLATION.
 RX MEDLINE=20556282; PubMed=10988299;
 RA Chi N.-W., Lodish H.F.;
 RT "Tankyrase is a Golgi-associated mitogen-activated protein kinase
 RT substrate that interacts with TRAP in GLUT4 vesicles";
 RL J. Biol. Chem. 275:38437-38444(2000).
 RN [4]

RP FUNCTION, AND MUTAGENESIS OF HIS-1184 AND GLU-1291.
 RA MEDLINE=21602874; PubMed=11739745;
 RX Cook B.D., Dynek J.N., Chang W., Shostak G., Smith S.;
 RT "Role for the related poly(ADP-Ribose) polymerases tankyrase 1 and 2
 RT at human telomeres";
 RL Mol. Cell. Biol. 22:332-342(2002).
 CC -1- FUNCTION: May regulate vesicle trafficking and modulate the
 CC subcellular distribution of SLC24/ GLUT4-vesicles. Has PARP
 CC activity and can modify TRF1, and thereby contribute to the
 CC regulation of telomere length.
 CC -1- CATALYTIC ACTIVITY: NAD(+) + {ADP-D-riboyl} (N)-acceptor =
 CC NICOTINAMIDE + [ADP-D-riboyl] (N+)-acceptor.
 CC -1- SUBUNIT: Oligomerizes and associates with TNKS2. Interacts with
 CC the cytoplasmic domain of LIMP2/ Oase in SLC24/ GLUT4-vesicles.
 CC Binds to the N-terminus of telomeric TRF1 via the ANK repeats.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic; associated with the Golgi and
 CC with juxtanuclear SLC24/ GLUT4-vesicles. A minor proportion is
 CC also found at nuclear pore complexes and around the pericentriolar
 CC matrix of mitotic centrosomes. During interphase, a small fraction
 CC of TNKS is found in the nucleus, associated with TRF1.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Name=1:
 CC IsoId=O95271-1; Sequence=Displayed;
 CC Name=2:
 CC IsoId=O95271-2; Sequence=VSP 004539, VSP 004539;
 CC Note=No experimental confirmation available;
 CC -1- TISSUE SPECIFICITY: Ubiquitous; highest levels in testis.
 CC -1- PTM: Upon insulin-stimulation, phosphorylated on serine residues
 CC by MARK kinases.
 CC -1- PTM: ADP-ribosylated (-auto).
 CC -1- SIMILARITY: Belongs to the PARP family.
 CC -1- SIMILARITY: Contains 15 ANK repeats.
 CC -1- SIMILARITY: Contains 1 sterile alpha motif (SAM) domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF082556; AAC79841.1; -
 DR EMBL; AF082557; AAC79842.1; -
 DR EMBL; AF082558; AAC79843.1; -
 DR EMBL; AF082559; AAC79844.1; -
 DR HSSP; Q00420; IAWC.
 DR GeneW; HGNC:11941; TNKS.
 DR MIM; 603303; -
 DR GO; GO:0000781; C:chromosome, telomeric region; IDA.
 DR GO; GO:0003950; F:NAD ADP-riboseyltransferase activity; IDA.
 DR GO; GO:0005515; F:protein binding; IDA.
 DR GO; GO:0007004; F:telomerase-dependent telomere maintenance; NAS.
 DR InterPro; IPR002110; ANK.
 DR InterPro; IPR001660; SAM.
 DR Pfam; PF00023; ank; 19.
 DR Pfam; PF00536; SAM; 1.
 DR PRINTS; PR01415; ANKYRIN.
 DR SMART; SM00248; ANK; 17.
 DR SMART; SM00454; SAM; 1.
 DR PROSITE; PS50088; ANK REPEAT; 15.
 DR PROSITE; PS50297; ANK REP REGION; 1.
 DR PROSITE; PS50105; SAM DOMAIN; 1.
 KW Transferase; Glycosyltransferase; NAD; Golgi stack; Telomere;
 KW Nuclear protein; Repeat; ANK repeat; ADP-riboylation;
 KW Phosphorylation; Alternative splicing.
 FT REPEAT 215 247 ANK 1.
 FT REPEAT 248 280 ANK 2.
 FT REPEAT 281 313 ANK 3.
 FT REPEAT 358 400 ANK 4.
 FT REPEAT 401 433 ANK 5.
 FT REPEAT 434 466 ANK 6.

```

FT REPEAT 521 556 ANK 7.
FT REPEAT 557 589 ANK 8.
FT REPEAT 590 622 ANK 9.
FT REPEAT 683 715 ANK 10.
FT REPEAT 716 748 ANK 11.
FT REPEAT 749 781 ANK 12.
FT REPEAT 836 868 ANK 13.
FT REPEAT 869 901 ANK 14.
FT REPEAT 902 934 ANK 15.
FT DOMAIN 1030 1089 SAM.
FT DOMAIN 1176 1327 PAMP.
FT DOMAIN 9 14 POLY-HIS.
FT DOMAIN 27 34 POLY-PRO.
FT DOMAIN 128 134 POLY-SER.
FT DOMAIN 137 145 POLY-SER.
FT VARSPLIC 641 643 EST -> GHS (in isoform 2).
FT VARSPLIC 644 1327 /FTID=VSP_004538.
FT VARSPLIC 644 1327 Missing (in isoform 2).
FT MUTAGEN 1184 1184 /FTID=VSP_004539.
FT MUTAGEN 1291 1291 H->A: LOSS OF ACTIVITY; WHEN ASSOCIATED WITH A-1291.
FT MUTAGEN 1291 1291 E->A: LOSS OF ACTIVITY; WHEN ASSOCIATED WITH A-1184.
SQ SEQUENCE 1327 AA; 142010 MW; E14DE985C710B957 CRC64;

Query Match 11.9%; Score 92; DB 1; Length 1327;
Best Local Similarity 30.4%; Pred. No. 3.3;
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

QY 6 VAAFAAI-VVSGSALAGVPPWGGGNNHNGSGSPDSTLSTIYQGSANALALQSDAR 64
DB 99 VAAAPVPAVSTSSAAGVAPNPAAGSGNSPSSSSPTSS-SSSSPSSPGSLAESPEAA 157
QY 65 KSEETIT-OSGYNGADVCGADNSTIELTONG-FRNATIDQWNAKNSDI 113
DB 158 GVSSTAPLPGCAAGGTGVPVAVGALRELLFACRGDVSRYKRLVDANVAKM 212

RESULT 7
OVO DROME
ID_OVO DROME STANDARD; PRT; 1028 AA.
AC P1521; Q9XZU4;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE OVO protein (Shaven baby protein).
GN OVO OR SVB.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; OC Ephydroidea; Drosophilidae; Drosophila.
OC NCBI_TaxID=7227;
RN [1]
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovary;
RA MEDLINE=95021209; PubMed=7935398;
RA Gadfinckel M.D., Wang J., Liang Y., Mahowald A.P.;
RT "Multiple products from the shavenbaby-ovo gene region of Drosophila melanogaster: relationship to genetic complexity.";
RL Mol. Cell. Biol. 14:6809-6818(1994).
RL [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Oregon-R;
RX MEDLINE=91293102; PubMed=1712294;
RA Mevel Nindo M.T.M., Terracol R., Kafatos F.C.;
RT "The ovo gene of Drosophila encodes a zinc finger protein required for female germ line development.";
RL EMBO J. 10:2259-2266(1991).
CC -1- FUNCTION: REQUIRED FOR SURVIVAL AND DIFFERENTIATION OF FEMALE GERM LINE CELLS. PLAYS A ROLE IN GERM LINE SEX DETERMINATION.
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -1- DEVELOPMENTAL STAGE: FIRST APPEARS IN THE GERMARINUM AND ACCUMULATES IN NURSE CELLS DURING OOGENESIS. STORED IN THE EGG,

```

```

CC BUT IS RAPIDLY LOST IN THE EMBRYOS EXCEPT FOR ITS CONTINUED
CC PRESENCE IN THE GERM LINE PRECURSOR POLE CELLS.
CC -1- SIMILARITY: Contains 4 C2H2-type zinc fingers.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-stb.ch/announce/
CC or send an email to license@isb-stb.ch).
CC -----
DR EMBL; U11383; AAB60216.1; -.
DR EMBL; X59772; CAB36921.1; ALT_SEQ.
DR PIR; A56038; A56038.
DR HSSP; P07248; 2ADR.
DR TRANSFAC; T00669; -.
DR FlyBase; FBgn0003028; ovo.
DR InterPro; IPR007087; ZnF_C2H2.
DR Pfam; PF00096; ZnF_C2H2; 3.
DR SMART; SM00355; ZnF_C2H2; 4.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 3.
DR KZ; Zinc-finger; Metal-binding; DNA-binding; Repeat; Nuclear protein;
KW Transcription regulation.
FT DOMAIN 62 66 POLY-ALA.
FT DOMAIN 72 77 POLY-GLY.
FT DOMAIN 80 85 POLY-GLY.
FT DOMAIN 98 108 POLY-GLY.
FT DOMAIN 144 152 POLY-HIS.
FT DOMAIN 153 159 POLY-ASN.
FT DOMAIN 336 339 POLY-GLN.
FT DOMAIN 347 353 POLY-GLN.
FT DOMAIN 357 361 POLY-GLN.
FT DOMAIN 410 414 POLY-GLN.
FT DOMAIN 418 422 POLY-GLN.
FT DOMAIN 426 432 POLY-GLN.
FT DOMAIN 445 453 POLY-GLN.
FT DOMAIN 456 459 POLY-GLN.
FT DOMAIN 466 474 POLY-GLN.
FT DOMAIN 497 517 POLY-ALA.
FT DOMAIN 524 529 POLY-SER.
FT DOMAIN 549 558 POLY-ALA.
FT DOMAIN 639 651 POLY-ALA.
FT DOMAIN 717 725 POLY-ALA.
FT DOMAIN 797 802 POLY-GLN.
FT DOMAIN 820 823 POLY-GLN.
FT DOMAIN 826 832 POLY-GLN.
FT ZN_FING 874 896 C2H2-TYPE 1.
FT ZN_FING 902 924 C2H2-TYPE 2.
FT ZN_FING 930 953 C2H2-TYPE 3.
FT ZN_FING 969 992 C2H2-TYPE 4.
FT CONFLICT 647 647 A->R (in Ref. 2).
SQ SEQUENCE 1028 AA; 110620 MW; D706BBB2BC0F6F77 CRC64;

Query Match 11.7%; Score 91; DB 1; Length 1028;
Best Local Similarity 31.2%; Pred. No. 3;
Matches 34; Conservative 31; Mismatches 41; Indels 26; Gaps 4;

QY 3 LKVAFAAI-VVSGSALAGVPPWGGGNNHNGSGSPDSTLSTIYQGSANALALQSD 62
DB 59 LQNAAAAIYMSAGS-----GGGCTGNGGGGSGGPGPSANSGGGGGG----- 104
QY 63 ARKSEETITOSGYNGADVCGADNSTIELTONGFRNATIDQWNAKNS 111
DB 105 -----GGNGYINGGVG-GPNNSS---LDGNLNLNPAVSNNVMSNS 141

RESULT 8
CSGB_SALTI STANDARD; PRT; 151 AA.
ID CSGB_SALTI
AC 0827M3;
DT 28-FEB-2003 (Rel. 41, Created)

```

DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Minor curlin subunit precursor.
 GN CSGB OR STY1180 OR T1777.
 OS Salmonella typhi.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 OC NCBI_TaxID=601;
 RX SEQUENCE FROM N.A.
 RC STRAIN=CT18;
 RX MEDLINE=21534947; PubMed=11677608;
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
 RA Cherkas S., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
 RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
 RA Kohn A., Larsen T.S., Leach S., Mouton R., O'Gea P., Parry C.,
 RA Quail M.A., Rutherford K., Simmonds M., Skellern J., Stevens K.,
 RA Whitehead S., Barrett B.G.;
 RT "Complete genome sequence of a multiple drug resistant Salmonella
 RT enterica serovar Typhi CT18.";
 RL Nature 413:848-852(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Ty2 / ATCC 700931;
 RX MEDLINE=22531367; PubMed=12644504;
 RA Deng W., Liou S.-R., Plunkett G., Iii, Mayhew G.F., Rose D.J.,
 RA Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;
 RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
 RT and CT18.";
 RL J. Bacteriol. 185:2330-2337(2003).
 CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
 CC CURLIN MONOMERS.
 CC -1- SIMILARITY: BELONGS TO THE CSGB/CSGB FAMILY.
 CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: A1627269; CAD08267.1; -
 DR EMBL: A6016840; AAC09400.1; -
 KW Fimbria; Signal; Complete proteome.
 FT SIGNAL 1 21
 FT CHAIN 22 151 MINOR CURLIN SUBUNIT.
 SQ SEQUENCE 151 AA; 16254 MW; 161C54326E573495 CRC4;
 Query Match 11.7%; Score 90.5; DB 1; Length 151;
 Best Local Similarity 25.4%; Pred. No. 0.36;
 Matches 33; Conservative 17; Mismatches 43; Indels 37; Gaps 4;
 QY 10 AAIVGSGALAGVPPWGGGNGGNSGPDSTLSIYGSANMALQSDAKSETT 69
 DB 58 AARROGSKLSTVISO--EGENNRARVDQGNVNFVIEPTGNN-----DAS 103
 QY 70 ITGSGGNGADVGADNSTIETLNGFRNATIDQNNANQSDITVQYDQLVTRVYTHE 129
 DB 104 IGSATGNSA-----AIQKSGSKAKITVGYGK-----TAVVQK 140
 QY 130 MAHASVVRQ 139
 DB 141 OSHMAIRVQ 150
 RESULT 9
 PER_DROWI

ID PER DROWI STANDARD: PRT; 1093 AA.
 AC 003297; 018421; 018422; P91721; P91722;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DE 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Period circadian protein (Fragment).
 GN PER.
 OS Drosophila willistonii (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OC NCBI_TaxID=7260;
 RX SEQUENCE FROM N.A.
 RC STRAIN=Various strains;
 RX MEDLINE=97357421; PubMed=9214747;
 RA Gleason J.M., Powell J.R.;
 RT "Interspecific and intraspecific comparisons of the period locus in
 RT the Drosophila willistonii sibling species.";
 RL Mol. Biol. Evol. 14:741-753(1997).
 RN [2]
 RP SEQUENCE OF 579-646 FROM N.A.
 RX MEDLINE=93196482; PubMed=8450754;
 RA Peixoto A.A., Campos S., Costa R.H., Kyriacou C.P.;
 RT "Molecular evolution of a repetitive region within the per gene of
 RT Drosophila.";
 RL Mol. Biol. Evol. 10:127-139(1993).
 CC -1- FUNCTION: Essential for biological clock functions. Determines the
 CC period length of circadian and ultradian rhythms; an increase in
 CC per dosage leads to shortened circadian rhythms and a decrease in
 CC leads to lengthened circadian rhythms. Essential for the circadian
 CC rhythmicity of locomotor activity, eclosion behavior, and for the
 CC rhythmic component of the male courtship song that originates in the
 CC the thoracic nervous system. The biological cycle depends on the
 CC rhythmic formation and nuclear localization of the TIM-PER
 CC complex. Light induces the degradation of TIM, which promotes
 CC elimination of PER. Nuclear activity of the heterodimer
 CC coordinately regulates PER and TIM transcription through a
 CC negative feedback loop. Behaves as a negative element in circadian
 CC transcriptional loop. Does not appear to bind DNA, suggesting
 CC indirect transcriptional inhibition (By similarity).
 CC -1- SUBUNIT: Forms heterodimer with timeless (TIM); the complex then
 CC translocates into the nucleus (By similarity).
 CC -1- SUBCELLULAR LOCATION: Nuclear at specific periods of the day.
 CC First accumulates in the perinuclear region about one hour before
 CC translocation into the nucleus. Interaction with Tim is required
 CC for nuclear localization (By similarity).
 CC -1- PTM: PHOSPHORYLATED WITH A CIRCADIAN RHYTHMICITY, PROBABLY BY THE
 CC DOUBLE-TIME PROTEIN (DBT). PHOSPHORYLATION COULD BE IMPLICATED IN
 CC THE STABILITY OF PER MONOMER AND IN THE FORMATION OF HETERODIMER
 CC PER-TIM (BY SIMILARITY).
 CC -1- MISCELLANEOUS: THE SEQUENCE SHOWN IS THAT OF STRAIN ATLIXCO.
 CC -1- SIMILARITY: Contains 1 PAS-associated C-terminal (PAS) domain.
 CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: U51055; AAB41360.1; -
 DR EMBL: U51056; AAB41361.1; -
 DR EMBL: U51057; AAB41362.1; -
 DR EMBL: U51058; AAB41363.1; -
 DR EMBL: U51059; AAB41364.1; -
 DR EMBL: U51060; AAB41365.1; -
 DR EMBL: U51061; AAB41366.1; -
 DR EMBL: U51062; AAB41367.1; -
 DR EMBL: U51063; AAB41368.1; -
 DR EMBL: U51064; AAB41369.1; -

```

OS Bacteriophage T2.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
CC T4-like viruses.
XX NCBI_TaxID=10664;
RM [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=87283911; PubMed=3302276;
RT "RNA sequence of gene 38 encoding a receptor-recognizing protein of bacteriophages T2, K3 and of K3 host range mutants."
RL J. Mol. Biol. 194:31-39(1987).
CC -!- FUNCTION: Vg38 is at the tip of the long tail fibers and serves as the phage recognition site for the cellular receptor.
CC -!- MISCELLANEOUS: THIS PHAGE USE OUTER MEMBRANE PROTEINS OMPF AND TTR AS RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.ebi.ac.uk/announcements or send an email to license@ebi-sib.ch).
CC
CC EMBL: X05312; CAA28935.1; -.
DR PIR: S00275; S00275.
DR InterPro: IPR007932; Tail_fibre_GP38.
DR Pfam: PF05268; GP38. 1.
SQ Fiber protein; Phage recognition.
SEQUENCE 262 AA; 25801 MW; 0567366918FcC745 CRC64;
Query Match 11.5%; Score 89.5; DB 1; Length 262;
Best Local Similarity 34.1%; Pred. No. 0.83;
Matches 30; Conservative 8; Mismatches 33; Indels 17; Gaps 4;
QY 27 GGGGNHNGGSSGPDSLTSTLYGGSAMALALSDARKSEKTTITSGYGNDGVQGD 86
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
DB 17S GGGGPFPGVGKGITGSDSLT-----GSMASL--TDAGTGTFV-QYGANGNVAGGG 225
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
QY 87 NSTIELTONGFRNATIDOWNAKNSDIT 114
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
DB 226 -----RGWKKNVITSBEGGAAGAAYT 245
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |

RESULT 11
PTIC_ERWCH STANDARD; PRT; 479 AA.
AC PI6317;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Secreted protease C precursor (BC 3.4.24.-) (Proc).
GC PTIC.
OS Erwinia chrysanthemi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pectobacterium.
OX NCBI_TaxID=556;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B374;
RX MEDLINE=91009140; PubMed=2211614;
RA Delapeltre P., Wandersman C.;
RT "Protein secretion in Gram-negative bacteria. The extracellular metalloprotease B from Erwinia chrysanthemi contains a C-terminal secretion signal analogous to that of Escherichia coli alpha-hemolysin."
RL J. Biol. Chem. 265:17118-17125(1990).
RN [2]
RP SEQUENCE OF 1-59 FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=B374;
RX MEDLINE=89255387; PubMed=2722818;
RA Delapeltre P., Wandersman C.;
RT "Protease secretion by Erwinia chrysanthemi. Proteases B and C are

```


RT synthesized and secreted as zymogens without a signal peptide.";
 RL J. Biol. Chem. 264:9083-9089(1989).
 CC -1- COFACTOR: Binds 1 zinc ion and 7 calcium ions per subunit (By similarity).
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- DOMAIN: THE GLY-RICH REPEATS MAY BE IMPORTANT IN THE EXTRACELLULAR SECRETION OF THIS METALLOPROTEASE.
 CC -1- SIMILARITY: Belongs to peptidase family M10B.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/announce/](http://www.isb-sib.ch/announce/or send an email to license@isb-sib.ch) or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL, MS9229; AAA24860.1; -;
 DR EMBL, J04736; AAA24862.1; -;
 DR EMBL, M60395; AAA63638.1; -;
 DR PIR, A38307; A38307.
 DR PDB, 1G07; 17-OCT-02.
 DR PDB, 1G08; 17-OCT-02.
 DR PDB, 1K7G; 19-OCT-02.
 DR PDB, 1K7Q; 20-OCT-02.
 DR MEROPS, M10.054; -;
 DR InterPro, IPR001343; Hemlysn_Ca_bind.
 DR InterPro, IPR006025; Pept_M_Zn_BS.
 DR InterPro, IPR006026; Peptidase_M.
 DR Pfam, PF00353; hemelysincabind; 3.
 DR PRINTS, PR00313; CABNDNGRPT.
 DR SMART, SM00235; ZMNC; 1.
 DR PROSITE, PS00142; ZINC_PROTEASE; 1.
 DR PROSITE, PS00330; HEMOLYSIN_CALCITUM; 1.
 DR Hydroxylase, Metalloprotease, Calcium-binding; Metal-binding; Zinc; Repeat; Zymogen; 3D-structure.
 KM Repeat; Zymogen; 3D-structure.
 FT PROPEP 1 17
 FT CHAIN 18 479
 FT REPEAT 345 353
 FT REPEAT 354 362
 FT REPEAT 363 371
 FT REPEAT 372 380
 FT REPEAT 381 389
 FT METAL 188 188
 FT ACT SITE 189 189
 FT METAL 192 192
 FT METAL 228 228
 FT METAL 265 265
 FT METAL 267 267
 FT METAL 297 297
 FT METAL 299 299
 FT METAL 300 300
 FT METAL 302 302
 FT METAL 339 339
 FT METAL 341 341
 FT METAL 346 346
 FT METAL 348 348
 FT METAL 350 350
 FT METAL 355 355
 FT METAL 357 357
 FT METAL 359 359
 FT METAL 363 363
 FT METAL 364 364

FT METAL 365 365
 FT METAL 366 366
 FT METAL 368 368
 FT METAL 372 372
 FT METAL 373 373
 FT METAL 375 375
 FT METAL 377 377
 FT METAL 381 381
 FT METAL 382 382
 FT METAL 383 383
 FT METAL 384 384
 FT METAL 386 386
 FT METAL 395 395
 FT METAL 402 402
 FT METAL 412 412
 SQ SEQUENCE 479 AA; 51600 MW; 990ED8376725DF61 CRC64;
 Query Match 11.4%; Score 88.5; DB 1; Length 479;
 Best Local Similarity 28.8%; Pred. No. 2;
 Matches 40; Conservative 21; Mismatches 41; Indels 37; Gaps 9;
 QY 27 GGGGNNNGGNSGSDPTSTIYVGSANALALQSDARSETTITQSG-----YNGG--- 78
 DB 345 GGGGNDILVGNNSD-----NIIQGGAGNDVLVGGAGA---DTLYGAGGRDTFYGGGQDS 396
 QY 79 -----ADVGGADNSTIELTQNGFRNNAT-----DQMANNSITVQGYD--QLVTR 124
 DB 397 TVAAAYDWIADPFGKID----KIDLSAFRNEGQLSFVQDQFTGKGQVWL-QWDANSITN 451
 QY 125 VVTHEMAHASY--MVRQVG 141
 DB 452 LMIHEAGHSSVDFVLAIVG 470
 RESULT 12
 GP63 LEIEME STANDARD; PRT; 646 AA.
 AC P43150;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Leishmanolysin Cl precursor (BC 3.4.24.36) (Cell surface protease)
 DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface endopeptidase).
 GN GP63-CI.
 OS Leishmania mexicana.
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 OX NCBI_Taxid=5665;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MNYC/Bz/62/M379;
 RX MEDLINE=93149206; PubMed=8426614;
 RA Medina-Acosta E., Karses R.E., Russell D.G.;
 RT "Structurally distinct genes for the surface protease of Leishmania mexicana are developmentally regulated.";
 RL Mol. Biochem. Parasitol. 57:31-46(1993).
 CC -1- FUNCTION: Has an integral role during the infection of macrophages in the mammalian host.
 CC -1- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and P1', and basic residues at P2 and P3'. A model nonapeptide is cleaved at -Ala-Tyr|-Leu-Lys-Lys--.
 CC -1- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
 CC -1- DEVELOPMENTAL STAGE: Expressed in both the promastigote and the

CC amastigote forms.
 CC -1- SIMILARITY: Belongs to peptidase family M8.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL: X64394; CAA45733.1; -
 DR PIR: S19916; S19916.
 DR HSSP: P08148; 1LML.
 DR MEROPS: M08.001; -
 DR GLYCOsuiteDB: P43150; -
 DR InterPro: IPR006025; Pept_M_Zn_BS.
 DR InterPro: IPR001577; Peptidase_M8.
 DR Pfam: PF01457; Peptidase_M8; 1.
 DR PRINTS: PS00782; LSHMANOLYSIN.
 DR PROSITE: PS00142; ZINC_PROTEASE; 1.
 DR Hydroxylase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;
 KW Zymogen; Signal; Cell adhesion; Multigene family.
 FT SIGNAL 1 39
 FT PROPE 40 102
 FT CHAIN 103 646
 FT METAL 266 266
 FT ACT_SITE 267 267
 FT METAL 270 270
 FT METAL 336 336
 FT DISULFID 127 144
 FT DISULFID 133 232
 FT DISULFID 316 388
 FT DISULFID 395 458
 FT DISULFID 408 427
 FT DISULFID 417 492
 FT DISULFID 469 513
 FT DISULFID 518 568
 FT DISULFID 538 561
 FT CARBOHYD 86 297
 FT CARBOHYD 297 399
 FT CARBOHYD 409 409
 FT CARBOHYD 433 433
 FT CARBOHYD 445 445
 FT CARBOHYD 466 466
 FT CARBOHYD 501 501
 FT SEQUENCE 646 AA; 69054 MW; FE44BDC78C10B0A CRC64;
 SQ
 Query Match 11.2%; Score 87; DB 1; Length 646;
 Best Local Similarity 73.1%; Pred. No. 3.9;
 Matches 19; Conservative 1; Mismatches 0; Indels 6; Gaps 1;
 QY 117 QYDQLVTRVVTHEMAHSAVMVROYGF 142
 Db 255 RYDQLVTRVVTHEMAHA-----VGF 274
 RESULT 13
 YF48 MYCTU STANDARD; PRT; 678 AA.
 AC 010778;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Hypothetical PPE-family protein Rv1548c/MT1599.
 GN Rv1548C OR MT1599 OR MTCY48.17.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 CX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=H37Rv;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekle A.F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Kirogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 RL Nature 393:537-544 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RX MEDLINE=22206494; PubMed=12218036;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
 RA Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
 RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
 RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains."
 RL J. Bacteriol. 184:5479-5490 (2002).
 CC -1- SIMILARITY: Belongs to the mycobacterial PPE family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL: Z74020; CAA98335.1; -
 DR EMBL: AE007026; AKA45866.1; ALT_INIT.
 DR PIR: A70762; A70762.
 DR TIGR: MT1559; -
 DR Tuberculin; Rv1548c; -
 DR InterPro: IPR00030; Microbac_PPE.
 DR InterPro: IPR002989; Mycobac_Pentapep.
 DR Pfam: PF01469; Pentapeptide_2; 11.
 DR Pfam: PF00823; PPE; 1.
 KW Hypothetical protein; Transmembrane; Complete proteome.
 FT TRANSMEM 14 34
 FT TRANSMEM 180 200
 FT TRANSMEM 258 258
 FT CONFLICT 258
 FT SEQUENCE 678 AA; 66736 MW; 209F1593D5253A2 CRC64;
 SQ
 Query Match 11.2%; Score 87; DB 1; Length 678;
 Best Local Similarity 26.2%; Pred. No. 4.1;
 Matches 34; Conservative 10; Mismatches 40; Indels 46; Gaps 7;
 QY 26 WGGG--GNHN--GGNNSGGPDTLSIYOGSANAALALQSPARKSETTITSGYGADVG 82
 Db 267 WGSNGIGSYNLGGN-----LGSYNLGSGN-----TGDITFGSGNTGNTLN 306
 QY 83 OGADNSTIELTON--GFRNATIDQWAKKSDITVQCYDQLVTRVVTHEMAHSAVMVRO 139
 Db 307 VGGG-----TGNNSFGGNTGNVFNFGNGTGTDTNFGS-----GNLGGN 346
 QY 140 VGFGNATYAN 149
 Db 347 IGFNGKSGSHN 356
 RESULT 14
 GP63 LEICH STANDARD; PRT; 599 AA.
 AC P15706;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Leishmanolysin precursor (EC 3.4.24.36) (Cell surface protease)
 GN (Major surface glycoprotein) (GP63 protein) (Promastigote surface
 endopeptidase).
 OS GP63.
 OC Leishmania chagasi.
 OC Eukaryota; Eukaryozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 NCBI_TaxID=44271;
 RX SEQUENCE FROM N.A.
 RP MEDLINE=90205976; PubMed=2320059;
 RT "Miller R.A., Reed S.G., Parsons M.;
 RT "Leishmania gp63 molecule implicated in cellular adhesion lacks an
 RT Arg-Gly-Asp sequence."
 RL Mol. Biochem. Parasitol. 39:267-274(1990).
 RP [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92112918; PubMed=1370484;
 RA Ramamoorthy R., Donelson J.E., Paetz K.E., Maybodi M., Roberts S.C.,
 RA Wilson M.E.;
 RT "Three distinct RNAs for the surface protease gp63 are differentially
 RT expressed during development of Leishmania donovani chagasi
 RT promastigotes to an infectious form."
 RL J. Biol. Chem. 267:1868-1895(1992).
 CC -1- FUNCTION: Has an integral role during the infection of macrophages
 CC in the mammalian host.
 CC -1- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and
 CC P1' and basic residues at P2 and P3'. A model nonapeptide is
 CC cleaved at -Ala-Tyr-Lys-Lys-
 CC -1- CORPACTOR: Binds 1 zinc ion per subunit (by similarity).
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
 CC -1- SIMILARITY: Belongs to peptidase family M8.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M80672; AAA29238.1; -;
 DR EMBL: M28527; AAA29235.1; -;
 DR PIR: A44951; A44951.
 DR HSSP: P08148; 1LML.
 DR MEROPS: M08.001; -;
 DR InterPro: IPR006025; Pept_M_Zn_BS.
 DR InterPro: IPR001577; Peptidase_M8.
 DR Pfam: PF01457; Peptidase_M8; 1.
 DR PRINTS: PR00782; LSHMANOLYSIN.
 DR PROSITE: PS00142; ZINC_PROTEASE; 1.
 KM Hydroxylase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;
 KM Hydrogen; Signal; Cell adhesion; GPI-anchor; Lipoprotein.
 FT SIGNAL 1 39
 FT PROPEP 40 97
 FT CHAIN 98 574
 FT PROPEP 575 599
 FT METAL 261 261
 FT ACT_SITE 262 262
 FT METAL 265 265
 FT METAL 331 331
 FT DISULFID 122 139
 FT DISULFID 188 227
 FT DISULFID 311 383
 FT DISULFID 390 452
 FT DISULFID 403 422
 FT DISULFID 412 486
 FT DISULFID 463 507
 FT DISULFID 512 562
 FT DISULFID 532 555
 FT CARBOHYD 297 297
 FT CARBOHYD 394 394
 FT LIPID 574 574

FT SQ SEQUENCE 599 AA; 63848 MW; 746730AE8E2A257C CRC64;
 Query Match 11.1%; Score 86; DB 1; Length 599;
 Best local Similarity 66.7%; Pred. No. 4.3;
 Matches 20; Conservative 2; Mismatches 6; Indels 2; Gaps 1;
 QY 117 QYDQLVTRVTHMAMASVVRQVGFQGNNA 146
 Db 250 RYDQLVTRVTHMAMAA--LGRSVGFQEGA 277
 RESULT 15
 AGOI ARATH STANDARD; PRT; 1048 AA.
 AC 004379; Q9LPE83;
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Argonaute protein.
 GN AGOI OR ARG48410 OR F11A17.3 OR T1N15.2.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia; TIGR=leaf;
 RX MEDLINE=98090460; PubMed=9427751;
 RA Bohmert K., Camus I., Bellini C., Bouchez D., Caboche M., Benning C.;
 RT "AGO1 defines a novel locus of Arabidopsis controlling leaf
 RT development."
 RL EMBL J. 17:170-180(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=21016719; PubMed=11130712;
 RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
 RA White O., Alonso J., Altieri H., Araujo R., Bowman C.L., Brooker S.Y.,
 RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
 RA Chung M.K., Conn L., Conway A.B., Conway A.R., Cressy T.H., Dewar K.,
 RA Dunn P., Egu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,
 RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Hultzer L.,
 RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
 RA Kim C.J., Koo H.L., Kremetska I., Kutz D.B., Kwan A., Lam B.,
 RA Langin-Hooper S., Lee A., Lee J.M., Ienz C.A., Li J.H., Li Y.-P.,
 RA Lin X., Liu S.X., Liu Z.A., Lueros U.S., Maiti R., Marziani A.,
 RA Miltescher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
 RA Pal G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
 RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
 RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
 RA Unterback T., Van Aken S., Vayberg M., Vysotskaia V.S., Walker M.,
 RA Wu D., Yu G., Frazer C.M., Venter J.C., Davis R.W.;
 RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
 RT thaliana."
 RL Nature 408:816-820(2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=22954850; PubMed=14593172;
 RA Yamada K., Lim J., Dale J.M., Chen H., Shinn P., Palm C.J.,
 RA Southwick A.M., Wu H.C., Kim C.J., Nguyen M., Pham P.K., Cheuk R.F.,
 RA Karlin-Nemenn G., Liu S.X., Lam B., Sakano H., Wu T., Yu G.,
 RA Miranda M., Quach H.L., Tripp C.S., Deng J.M., Akiyama K., Ansari Y.,
 RA Arakawa T., Banh J., Banno F., Bowser L., Brookes S.Y., Carninci P.,
 RA Chao Q., Choy N., Enju A., Goldsmith A.D., Gurjal M., Hansen N.F.,
 RA Hayashizaki Y., Johnson-Hopson C., Heuan V.W., Iida K., Kamei M.,
 RA Khan S., Koesema B., Ishida J., Jiang P.X., Jones T., Kawai J.,
 RA Kamiya A., Meyers C., Nakajima M., Narusaka M., Seki M., Sakurai T.,
 RA Satou M., Tamari R., Vayberg M., Wallender E.K., Wong C., Yamamura Y.,
 RA Yuan S., Shinzaki K., Davis R.W., Theologis A., Ecker J.R.;

```
RT "Empirical analysis of transcriptional activity in the Arabidopsis
RL genome";
CC Science 302:842-846(2003).
CC -|- FUNCTION: Essential for proper development of leaves and floral
CC organs, and formation of axillary meristems.
CC -|- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -|- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=1;
CC Comment=A number of isoforms are produced. According to EST
CC sequences;
CC Name=1;
CC isoId=O04379-1; Sequence=Displayed;
CC -|- TISSUE SPECIFICITY: Widely expressed at low levels.
CC -|- DEVELOPMENTAL STAGE: Expressed throughout all developmental
CC stages.
CC -|- SIMILARITY: Belongs to the argonaute family.
CC -|- SIMILARITY: Contains 1 PAZ domain.
CC -|- SIMILARITY: Contains 1 Piwi domain.
CC -|- CAUTION: Ref.2 (AAF9718) sequence differs from that shown due to
CC erroneous gene model prediction.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U91995; AAC18440.1; -
DR EMBL: AC007932; AAD49755.1; -
DR EMBL: AC020889; AAF79718.1; ALT_SEQ.
DR EMBL: BT000941; AAN41341.1; -
DR InterPro: IPR003100; PAZ.
DR InterPro: IPR003165; Piwi.
DR Pfam: PF02170; PAZ; 1.
DR Pfam: PF02171; Piwi; 1.
DR PROSITE: PS50821; PAZ; 1.
DR PROSITE: PS50822; Piwi; 1.
DR Developmental protein; Alternative splicing.
KW DOMAIN 391 501 PAZ.
FT DOMAIN 676 997 Piwi.
FT DOMAIN 13 104 GLY-RICH.
SQ SEQUENCE 1048 AA; 116190 MW; 3E5146343A09C541 CRC64;

Query Match 11.0%; Score 85.5; DB 1; Length 1048;
Best Local Similarity 25.7%; Pred. No. 9.2;
Matches 36; Conservative 17; Mismatches 60; Indels 27; Gaps 3;

QY 27 GCGGNHNGGNSGPDSTLSI-----YQGSANMALALQSDARKSETTTTOSGYG 76
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 95 GGYGGGGGGSSGPPQGSVPELHQATSPYQAVSSQPTLSEVSPQVPEPTVLAQGF- 153
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 77 NGADVGGQADNSTLE-----LTQNGFRNNATTIDQWAKNSDITVGQYDQ 120
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 154 EQLSVEGQAPSOALQPTPSSSKAFKFPMPKQSGKRCIVKANHFPALPDKDLHHYDV 213
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 121 LVTRVVTHEMAHASVWVQV 140
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 214 TTFPEVTSRGVNRVAMKQLV 233
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Search completed: March 11, 2004, 18:34:54
Job time : 8.3 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using ew model

Run on: March 11, 2004, 18:23:54 ; Search time 30.5 Seconds
(without alignments)
1562.074 Million cell updates/sec

Title: US-09-543-407-14

Perfect score: 775

Sequence: 1 MLLKVAAPAAIVSGSALA.....HASVMVROVGFNNATANY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPREMBL_25.*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_humani:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	684	88.3	152	2	033802 salmonella
2	578.5	74.6	150	2	07X243 citrobacter
3	533	68.8	149	2	07X240 citrobacter
4	503.5	65.0	152	16	08CW63 escherichia
5	438.5	56.6	150	2	07X237 enterobacte
6	385	49.7	76	2	054069 salmonella
7	122	15.7	29	2	098335 escherichia
8	109	14.1	179	2	033801 salmonella
9	107.5	13.9	151	2	07X238 enterobacte
10	105	13.5	139	16	08E1H3 shewanella
11	104.5	13.5	152	2	07X241 citrobacter
12	104.5	13.5	160	16	08CW64 escherichia
13	104	13.4	502	16	08E1H4 shewanella
14	103.5	13.4	151	2	07X244 citrobacter
15	103.5	13.4	151	16	07UC21 shigella fl
16	103.5	13.4	160	16	083RU7 shigella fl

17	101	13.0	362	16	08EV84 mycoplasma
18	99.5	12.8	348	13	093397 cytrinus ca
19	96.5	12.5	552	16	P96840 mycobacteri
20	96.5	12.5	552	16	07TW76 mycobacteri
21	96.5	12.5	623	16	08VIY0 mycobacteri
22	96.5	12.5	1422	16	08EFU3 shewanella
23	96.5	12.5	3501	16	08Y106 raietonia s
24	96.5	12.5	3552	16	08XSD6 raietonia s
25	96	12.4	154	16	089J15 bradyrhizob
26	96	12.4	157	16	088HG0 pseudomonas
27	96	12.4	606	3	09P319 neurospora
28	95.5	12.3	624	3	08NIV1 neurospora
29	95.5	12.3	1765	16	07V8S5 prochlorococ
30	95	12.3	1748	5	094821 tetrahymena
31	94.5	12.2	191	3	08TF66 saccharomyc
32	94	12.1	171	16	089J13 bradyrhizob
33	94	12.1	368	16	08BMD6 mycoplasma
34	93	12.0	480	16	089EV2 bradyrhizob
35	93	12.0	2174	16	092U08 rhizobium m
36	92	11.9	453	5	09N6M8 drosophila
37	92	11.7	301	10	09X110 arabidopsis
38	91	11.7	453	5	09NGF6 drosophila
39	91	11.7	453	5	09NGF7 drosophila
40	91	11.7	1222	5	09M4F0 drosophila
41	91	11.7	1222	5	08T8L9 drosophila
42	91	11.7	1351	5	08SX56 drosophila
43	91	11.7	1354	5	08MPN4 drosophila
44	90.5	11.7	145	16	08U6N9 agrobacteri
45	90.5	11.7	191	10	07XDR3 oryza sativ

ALIGNMENTS

RESULT 1	033802	PRELIMINARY;	PRT;	152 AA.
ID	033802	01-JAN-1998 (TREMBLrel. 05, Created)		
AC	033802	01-JAN-1998 (TREMBLrel. 05, Last sequence update)		
DT	01-JAN-1998 (TREMBLrel. 05, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE	Agfa protein (Fragment).			
GN	Agfa.			
OS	Salmonella typhimurium.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OC	Enterobacteriaceae; Salmonella.			
OK	NCBI_Taxid=602;			
RN	(1)			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=98053981; PubMed=9393832;			
RA	Sukdolvi S.S., Lorentz R.G., Gordon J.I., Bian Z., Pfeiffer J.D.,			
RA	Normark S.J., Rhan M.;			
RT	"Expression of thin, aggregative fimbriae promotes interaction of			
RT	Salmonella typhimurium SR-11 with mouse small intestinal epithelial			
RT	cells."			
RL	Infect. Immun. 65:5320-5325(1997).			
DR	EMBL; AJ000514; CAA04151.1; -			
FT	NON_TER 152 152			
SQ	SEQUENCE 152 AA; 15401 MW; 9DA7DADC2364B006 CRC64;			
Query Match	88.3%; Score 684; DB 2; Length 152;			
Best Local Similarity	89.4%; Pred. No. 1.2e-47;			
Matches	135; Conservative			
Qy	1	MLLKVAAPAAIVSGSALAGVPPWGGGNNGGSSGPGSTSIYVGSANALAAQ	60	
Db	1	MLLKVAAPAAIVSGSAVAVPQWGGGNNGGSSGPGSTSIYVGSANALAAQ	60	
Qy	61	SDARKSETTITGSGNGADVCGAGDNSTIELTONGFNNATIDOMNANSDITVGOYQD	120	
Db	61	SDARKSETTITGSGNGADVCGAGDNSTIELTONGFNNATIDOMNANSDITVGOYQD	120	
Qy	121	LTVRVVTHMAHSAVMVROVGFNNATANY 151		

Db 121 NNAALVNOTASDSSVWVROVGFNNATNOY 151

RESULT 2

Q7X243 PRELIMINARY; PRT; 150 AA.
 ID 07X243;
 AC 07X243;
 DT 01-OCT-2003 (TREMBLrel. 25, Created)
 DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)
 DE CuriIn-csgA protein.
 GN CsgA.
 OS Citrobacter sp. Fec2.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Citrobacter.
 OX NCBI_TaxID=213763;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Fec2;
 RA Zogaj X., Bokranz W., Nimtz M., Romling U.;
 RT "Production of Cellulose and CuriI Fimbriae by Members of the Family Enterobacteriaceae Isolated from the Human Gastrointestinal Tract."
 RL Infect. Immun. 72:4151-4158(2003).
 DR EMBL: AJ515700; CAD56672.1;
 SQ SEQUENCE 150 AA; 15016 MW; 1D7141B8D6973DC6 CRC64;

Query Match 74.6%; Score 578.5; DB 2; Length 150;
 Best Local Similarity 76.8%; Pred. No. 3, 7e-39;
 Matches 116; Conservative 13; Mismatches 21; Indels 1; Gaps 1;

Qy 1 MKLTKVAAFAIIVSGSALAGVPPQWGGGNNNGSGSPDSTLSIYOGSANAALALQ 60
 Db 1 MKLTKVAAFAIIVSGSALAGVPPQWGGG-GGGSSSGSESTLSIYGVNNAALALQ 59
 Qy 61 SPARKSETTITGSGYNGADVQAGADNSTIELTONGFNNATTIQWNAKSDITVGOYD 120
 Db 60 SPARKSDTTIHONGFNGADVQGSNDSTIDLTQGFNNATTIQWNAKSDITVGOYD 119
 Qy 121 LVTRVVTHEMAHSAVWVROVGFNNATNOY 151
 Db 120 HNAALVNOTASDSSVWVROVGFNNATNOY 150

RESULT 3

Q7X240 PRELIMINARY; PRT; 149 AA.
 ID 07X240;
 AC 07X240;
 DT 01-OCT-2003 (TREMBLrel. 25, Created)
 DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)
 DE CuriIn-csgA protein.
 GN CsgA.
 OS Citrobacter freundii.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Citrobacter.
 OX NCBI_TaxID=546;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Fec4;
 RA Zogaj X., Bokranz W., Nimtz M., Romling U.;
 RT "Production of Cellulose and CuriI Fimbriae by Members of the Family Enterobacteriaceae Isolated from the Human Gastrointestinal Tract."
 RL Infect. Immun. 72:4151-4158(2003).
 DR EMBL: AJ515701; CAD56675.1;
 SQ SEQUENCE 149 AA; 15260 MW; 946DD52017F648FD CRC64;

Query Match 68.8%; Score 533; DB 2; Length 149;
 Best Local Similarity 70.9%; Pred. No. 1, 7e-35;
 Matches 107; Conservative 20; Mismatches 22; Indels 2; Gaps 1;

Qy 1 MKLTKVAAFAIIVSGSALAGVPPQWGGGNNNGSGSPDSTLSIYOGSANAALALQ 60
 Db 1 MKLTKVAAFAIIVSGSALAGVPPQWGGGNNNGSGSPDSTLSIYGVNNAALALQ 59

Db 1 MKLTKVAAFAIIVSGSALAGVPPQW--GGNHGGSGNYGPDSSLSIYOGSNNNSAALQ 58

Qy 61 SDARKSETTITGSGYNGADVQAGADNSTIELTONGFNNATTIQWNAKSDITVGOYD 120

Db 59 SDARKSDTTIHONGFNGADVQAGADNSTIELTQGFNNATTIQWNAKSDITVGOYD 118

Qy 121 LVTRVVTHEMAHSAVWVROVGFNNATNOY 151
 Db 119 RNGALVNOTASDSSVWVROVGFNNATNOH 149

RESULT 4

Q8CW63 PRELIMINARY; PRT; 152 AA.
 ID Q8CW63;
 AC Q8CW63;
 DT 01-MAR-2003 (TREMBLrel. 23, Created)
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
 DE Major curliin subunit precursor.
 GN CsgA OR C1306.
 OS Escherichia coli O6.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=217992;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O6:H1 / ATCC 700928;
 RX MEDLINE=22388234; PubMed=12471157;
 RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
 RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
 RT "Extensive mosaic structure revealed by the complete genome sequence of uropathogenic Escherichia coli."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
 DR EMBL: AE016759; AAN79779.1;
 KW Complete proteome.
 SQ SEQUENCE 152 AA; 15064 MW; 3BA57F34C1240E83 CRC64;

Query Match 65.0%; Score 503.5; DB 16; Length 152;
 Best Local Similarity 67.1%; Pred. No. 4, 1e-33;
 Matches 102; Conservative 20; Mismatches 29; Indels 1; Gaps 1;

Qy 1 MKLTKVAAFAIIVSGSALAGVPPQW--GGGNNNGSGSPDSTLSIYOGSANAALALQ 59
 Db 1 MKLTKVAAFAIIVSGSALAGVPPQYGGGNGGSGNNSGPSELTNIYOGGNSALALQ 60
 Qy 60 QSDARKSETTITGSGYNGADVQAGADNSTIELTONGFNNATTIQWNAKSDITVGOYD 119
 Db 61 QADARNSDLTITOHGGNGADVQGSDDSIDLTQGFNSATLIDQWNAKSDITVGOYD 120
 Qy 120 QLVTRVVTHEMAHSAVWVROVGFNNATNOY 151
 Db 121 GNGAALVNOTASNSVWVROVGFNNATNOY 152

RESULT 5

Q7X237 PRELIMINARY; PRT; 150 AA.
 ID 07X237;
 AC 07X237;
 DT 01-OCT-2003 (TREMBLrel. 25, Created)
 DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)
 DE CuriIn-csgA protein.
 GN CsgA.
 OS Enterobacter sakazakii.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Enterobacter.
 OX NCBI_TaxID=28111;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Fec39;
 RA Zogaj X., Bokranz W., Nimtz M., Romling U.;

Qy 1 MKLTKVAAFAIIVSGSALAGVPPQWGGGNNNGSGSPDSTLSIYOGSANAALALQ 60
 Db 1 MKLTKVAAFAIIVSGSALAGVPPQWGGGNNNGSGSPDSTLSIYGVNNAALALQ 59

```

RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RT Enterobacteriaceae isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158(2003).
SQ SEQUENCE 150 AA; 15112 MW; 5DBBB2D872DF15F3 CRC64;

Query Match
Best Local Similarity 56.6%; Score 438.5; DB 2; Length 150;
Matches 89; Conservative 28; Mismatches 33; Indels 1; Gaps 1;

QY 1 MLLKVAAPAAIVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYOGSANAALALO 60
DB 1 MRFIKVAAIAIVSSASAMAGTINQ-GGNGHGHGCGYGGPNSTLITVYGGNSALALO 59
QY 61 SDARKSETTITQSGYGNAGADVGGADNSTIELTONGFRNNATIDQNNAKNSDITVGYDQ 120
DB 60 TQARNSVLINSIQGGNGAGADVGGSDSSINLTONGFGSATILDQNSNDSVWNVQYGG 119
QY 121 LVTRVYTHMAIASVWVRQVGFNNATNQY 151
DB 120 LMGALVDOTASNSTVWVTQIGFGNATAHQY 150

RESULT 6
Q54069 PRELIMINARY; PRT; 76 AA.
AC Q54069;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE SEF17 fimbria (fragment).
GN AGFA.
OS Salmonella enteritidis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=592;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SE30;
RA Cox J.M., Eglezos S., Woolcock J.B.;
RT "Virulence of Salmonella enteritidis in chickens correlates with
RT colony morphology and expression of SEF17 fimbriae.";
RL Submitted (Apr-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U53207; AAA98671.1; -.
FT NON TER 1 76
FT 1
SQ SEQUENCE 76 AA; 7704 MW; 2FDS411241A7BCB1 CRC64;

Query Match
Best Local Similarity 49.7%; Score 385; DB 2; Length 76;
Matches 75; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 30 GHHNGGSSGPDSTLSIYOGSANAALALQSDARKSETTITQSGYGNAGADVGGADNST 89
DB 1 GHHKGGSSGPDSTLSIYOGSANAALALQSDARKSETTITQSGYGNAGADVGGADNST 60
QY 90 IELTONGFRNNATIDQ 105
DB 61 IELTONGFRNNATIDQ 76

RESULT 7
Q9S3J5 PRELIMINARY; PRT; 29 AA.
AC Q9S3J5;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DB Curliin subunit monomer (fragment).
GN CSGA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.

```

```

OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC TRANSPOSON-insertion sequence ISI;
RX MEDLINE=99314153; PubMed=10386375;
RA La Ragione R.M., Collighan R.J., Woodward M.J.;
RT "Non-clonality of Escherichia coli O78:H80 isolates associated with
RT ISI insert on in csgB and reduced persistence in poultry infection.";
RL FEMS Microbiol. Lett. 175:247-253(1999).
DR EMBL; AJ131756; CAB45380.1; -.
FT NON TER 29 29
SQ SEQUENCE 29 AA; 2789 MW; E29DPC07ABBB243 CRC64;

Query Match
Best Local Similarity 15.7%; Score 122; DB 2; Length 29;
Matches 26; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MLLKVAAPAAIVSGSALAGVVPQWGGG 29
DB 1 MLLKVAAPAAIVSGSALAGVVPQYGGG 29

RESULT 8
O33801 PRELIMINARY; PRT; 179 AA.
AC O33801;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE AgfB protein.
GN AGFB.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=98053981; PubMed=9393832;
RA Sukupolvi S.S., Lorentz R.G., Gordon J.I., Bian Z., Pfeifer J.D.,
RA Normark S.J., Rhen M.;
RT "Expression of thin, aggregative fimbriae promotes interaction of
RT Salmonella typhimurium SR-11 with mouse small intestinal epithelial
RT cells.";
RL Infect. Immun. 65:5320-5325(1997).
DR EMBL; AJ000514; CA04150.1; -.
SQ SEQUENCE 179 AA; 19318 MW; A2BCCB648B3C0B08 CRC64;

Query Match
Best Local Similarity 14.1%; Score 109; DB 2; Length 179;
Matches 36; Conservative 20; Mismatches 48; Indels 32; Gaps 4;

QY 16 GSALAGVVPQWGGGNNHGGNSGPDSTLSIYOGSANAALALQSDARKSETTITQSGY 75
DB 63 GSKLSTVLSQ-ERGANNAKVDQAGVNFAYIEQGNAN-----DASISQSN 109
QY 76 GNGADVGGADNSTIELTONGFRNNATIDQNNAKNSDITV-----GQYDQLV 123
DB 110 GNSAAI-----SAATIGKSGNKANITQYGTQTAVVVQKSHMAIQANITQYGTQ 162
QY 124 RYVTHMAIASVWVRQ 139
DB 163 AVVQKOSHMAIRVQ 178

RESULT 9
Q7X238 PRELIMINARY; PRT; 151 AA.
AC Q7X238;
DT 01-OCT-2003 (TREMBLrel. 25, Created)
DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Nucleation component of curliin monomers.
GN CSGB.

```

OS Enterobacter sakazakii.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Enterobacter.
 OX NCBI_TaxID=28141;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Ec39;
 RA Zogaj X., Bokranz W., Nintz M., Romling U.;
 RT "Production of Cellulose and Curli Fimbriae by Members of the Family
 Enterobacteriaceae Isolated from the Human Gastrointestinal Tract."
 RL Infect. Immun. 72:4151-4158(2003).
 DR EMBL: AJ515702; CAD56677.1;
 SQ SEQUENCE 151 AA; 15985 MW; F0B82BD2A2782B7 CRC64;

Query Match 13.9%; Score 107.5; DB 2; Length 151;
 Best Local Similarity 30.8%; Pred. No. 0.31;
 Matches 36; Conservative 14; Mismatches 42; Indels 25; Gaps 4;
 QY 10 AAIYVSSALAGVPPQWGGGNNHGGNSGPDSTLSIYOGSANAALAQDARKSETT 69
 DB 58 AQRQSSKLLSVSQ-----DGAGNPARVD-----QSGTYNIAWIDQS-GNGNDAG 103
 QY 70 ITOSGYNGADVCGADNSTIELTONGFRNATIDOMNANSDITVGYDQLVTRV 126
 DB 104 ITDQYGNNSA-----KTIQKSGNRANITGYGTQKTAVVVQKQSQMAIRVI 149

RESULT 10
 Q8ETH3
 ID O8ETH3 PRELIMINARY; PRT; 139 AA.
 AC O8ETH3;
 DT 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Minor curlin subunit CsgB, putative.
 GN S00866.
 OS Shewanella oneidensis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
 OC Alteromonadaceae; Shewanella.
 OX NCBI_TaxID=70863;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NR-1;
 RX MEDLINE=22297686; PubMed=12368813;
 RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaides E.J., Nelson W.C.,
 Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
 Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
 Debey R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
 Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
 Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,
 Mueller J., Khouri H., Gill J., Ullrich T.R., McDonald L.A.,
 Feldblyum T.V., Smith H.O., Venter J.C., Neilson K.H., Fraser C.M.,
 RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
 Shewanella oneidensis."
 RL Nat. Biotechnol. 20:1118-1123(2002).
 DR EMBL: AE015532; AAN53942.1;
 DR TIGR: S00866;
 KW Complete proteome.
 SQ SEQUENCE 139 AA; 14811 MW; 41EC1CFA76595720 CRC64;

Query Match 13.5%; Score 105; DB 16; Length 139;
 Best Local Similarity 28.3%; Pred. No. 0.45;
 Matches 32; Conservative 23; Mismatches 44; Indels 14; Gaps 4;
 QY 39 SGPDSSTSIYOGSANAALAQDARKSETTISYOGSANGAVGCGADNSTIELTONGFR 98
 DB 41 SGRDNILIDLVOQGTANGIYFGSGSNS-AIYVQAGNDNISLVTOGTNNEVQLQVGAQ 99
 QY 99 NNAITDQMAKNSDITVGYDQLVTRVTHEMAHSAVWRQVGFANATANQY 151
 DB 100 NKASITQ-----IGN-DNLVQ---LNLQSGNFSIQIADAAISITQY 139

RESULT 11
 Q7X241
 ID Q7X241 PRELIMINARY; PRT; 152 AA.
 AC Q7X241.
 DT 01-OCT-2003 (TREMBlrel. 25, Created)
 DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Nucleation component of curlin monomers.
 GN C5GB.
 OS Citrobacter freundii.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Citrobacter.
 OX NCBI_TaxID=546;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Fec3;
 RA Zogaj X., Bokranz W., Nintz M., Romling U.;
 RT "Production of Cellulose and Curli Fimbriae by Members of the Family
 Enterobacteriaceae Isolated from the Human Gastrointestinal Tract."
 RL Infect. Immun. 72:4151-4158(2003).
 DR EMBL: AJ515701; CAD56674.1;
 SQ SEQUENCE 152 AA; 16149 MW; D063A527D45D4329 CRC64;

Query Match 13.5%; Score 104.5; DB 2; Length 152;
 Best Local Similarity 29.1%; Pred. No. 0.55;
 Matches 32; Conservative 13; Mismatches 40; Indels 25; Gaps 3;
 QY 16 GSALGVPPQWGGGNNHGGNSGPDSTLSIYOGSANAALAQDARKSETTITQSGY 75
 DB 65 GSKLISVTSQ-----BGSNRAKTDP-----GSINPAY-----IDQTGS 99
 QY 76 GNGAVGCGADNSTIELTONGFRNATIDOMNANSDITVGYDQLVTRV 125
 DB 100 SMDASIKQSYGNTAVIITQKSGNKANITGYGTQKTAVVVQKQSQMAIRVI 149

RESULT 12
 O8CW64
 ID O8CW64 PRELIMINARY; PRT; 160 AA.
 AC O8CW64;
 DT 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Minor curlin subunit precursor.
 GN C5GB OR C1305.
 OS Escherichia coli O6.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=217992;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
 RX MEDLINE=22388234; PubMed=12471157;
 RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesech P.,
 Raeko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
 Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
 RA Mobley H.L.T., Domeneberg M.S., Blattner F.R.;
 RT "Extensive mosaic structure revealed by the complete genome sequence
 of uropathogenic Escherichia coli."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
 DR EMBL: AE016759; AAN79778.1;
 KW Complete proteome.
 SQ SEQUENCE 160 AA; 16963 MW; 49F68448D979B986 CRC64;

Query Match 13.5%; Score 104.5; DB 16; Length 160;
 Best Local Similarity 29.2%; Pred. No. 0.58;
 Matches 35; Conservative 19; Mismatches 59; Indels 7; Gaps 4;
 QY 9 PAIVVSSAL--AGVPPQWGGGNNHGGNSGPDSTLSIY--OYGSANAALAQDARK 65
 DB 42 FAVNELSKSFSFQAAI---GQAGTNNAGALRQSGSKLLITVADEGSSNRA-KIDQYGDY 97
 QY 66 SETTITQSGYNGADVCGADNSTIELTONGFRNATIDOMNANSDITVGYDQLVTRV 125

Db 98 NLAIVDQASANDASISQAGYNTAMIIQKSGNKNANITQYGTQKTAIVVQROSQMAIRV 157

RESULT 13

Q8EIH4 PRELIMINARY; PRT; 502 AA.

AC Q8EIH4; 01-MAR-2003 (TrEMBLrel. 23, Created)

DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)

DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

GN Conserved hypothetical protein.

OS S00865.

OC Shewanella oneidensis.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;

OC Alteromonadales; Shewanella.

OX NCBI_TaxId=70863;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=NR-1;

RX MEDLINE=22297686; PubMed=12368813;

RA Heidelberg J.F., Paulsen I.T., Nelson K.B., Gaidos E.J., Nelson W.C., Read T.D., Eisen J.A., Seshadri R., Ward N., McEhe B., Clayton R.A., Meyer R.T., Tsapin A., Scott J., Beaman M., Brinkac L., Daugherty S., DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F., Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M., Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C., Mueller J., Khouri H., Gill J., Uterback T.R., McDonald L.A., RA Feldlyum T.V., Smith H.O., Venter J.C., Nelson K.H., Fraser C.M.; RT "Genome sequence of the dissimilatory metal ion-reducing bacterium Shewanella oneidensis.";

RL Nat. Biotechnol. 20:1118-1123(2002).

DR EMBL; AE015532; AAN5941.1; -

RM TIGR; S00865; -

KV Hypothetical protein; Complete proteome.

SQ SEQUENCE 502 AA; 52441 MW; D08CA23D6C46862D CRC64;

Query Match 13.4%; Score 104; DB 16; Length 502; Best Local Similarity 26.0%; Pred. No. 2.5; Matches 40; Conservative 22; Mismatches 56; Indels 36; Gaps 7;

QY 29 GGNHNG-----GGN-----SGPDSLTLSIYQGSANA---ALALQS 61

Db 231 GNNHGFVYALAGSENDISMEQSGNNNTAYLSMTTGDPTVDITQDGSNTVDSIADI 290

QY 62 DAKSETTITQSGYGGADVQAGADNSTIELTONGFRNNTATIDQWAKXS-DI-TVGOYD 119

Db 291 QGDDNDITIKQKDSNGAGFQWVGSDNDVDLKRQGANPATRGAIVGTDDPDLSSKGDNN 350

QY 120 QLVTRVTHEMAHASYVRQVFGN-----NATAN 149

Db 351 ELVAFATGED---NSIEISQEGDANFAVYDATGN 381

RESULT 14

Q7X244 PRELIMINARY; PRT; 151 AA.

AC Q7X244; 01-OCT-2003 (TrEMBLrel. 25, Created)

DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)

DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

GN Nucleation component of curlin monomers.

OS CSGB.

OC Citrobacter sp. Fec2.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Citrobacter.

OX NCBI_TaxId=213763;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Fec2;

RA Zogaj X., Bokranz W., Nimtz M., Romling U.; RT "Production of Cellulose and Curli Fimbriae by Members of the Family Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";

RL Infect. Immun. 72:4151-4158(2003).

DR EMBL; A2515700; CAD56671.1; -

SQ SEQUENCE 151 AA; 16158 MW; BD00AF57E1400704 CRC64;

Query Match 13.4%; Score 103.5; DB 2; Length 151; Best Local Similarity 28.8%; Pred. No. 0.65; Matches 32; Conservative 13; Mismatches 41; Indels 25; Gaps 3;

QY 16 GSALAGVPQWGGGNGNNGGNSGSPDSTLSIYQGSANAALALQSDARKSETTITQSGY 75

Db 64 GSKLSVVSQ-----EGSNRAKVD-----QSAVYFAY-----IAQSH 98

QY 76 GNGADVQAGADNSTIELTONGFRNNTATIDQWAKNSDIVGYDQLVTRV 126

Db 99 SNDASISQSNYNTAMIIQKSGNKNANITQYGTQKTAIVVQROSQMAIRV 149

RESULT 15

Q7UC21 PRELIMINARY; PRT; 151 AA.

AC Q7UC21; 01-OCT-2003 (TrEMBLrel. 25, Created)

DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)

DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

GN Minor curlin subunit.

OS CSGB OR S1108.

OC Shigella flexneri.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Shigella.

OX NCBI_TaxId=623;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=2457T / ATCC 700930 / Serotype 2a;

RX MEDLINE=22590274; PubMed=12704152;

RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W., Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A., Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S., Schwartz D.C., Blattner F.R.; RT "Complete genome sequence and comparative genomics of Shigella flexneri serotype 2a strain 2457T.";

RL Infect. Immun. 71:2775-2786(2003).

DR EMBL; AB016981; AAP16542.1; -

SQ SEQUENCE 151 AA; 15868 MW; 5D5D266B964014A0 CRC64;

Query Match 13.4%; Score 103.5; DB 16; Length 151; Best Local Similarity 28.3%; Pred. No. 0.65; Matches 34; Conservative 20; Mismatches 59; Indels 7; Gaps 4;

QY 9 PAIVVSGAL--AGVPQWGGGNGNNGGNSGSPDSTLSIY-QYGSANAALALQSDARK 65

Db 33 PAVNELSKSSFNQAAII---GQAGTNNSSAQLRQGSKLLAAVVAQEGSSNRA-KIDQTCY 88

QY 66 SETTITQSGYGGADVQAGADNSTIELTONGFRNNTATIDQWAKNSDIVGYDQLVTRV 125

Db 89 NLAIVDQASANDASISQAGYNTAMIIQKSGNKNANITQYGTQKTAIVVQROSQMAIRV 148

Search completed: March 11, 2004, 18:40:12

Job time : 31.5 secs

THIS PAGE BLANK (USPTO)

CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell) the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

Sequence 151 AA;

Query Match 100.0%; Score 757; DB 3; Length 151;
 Best Local Similarity 100.0%; Pred. No. 2.2e-67;
 Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLKVAAFAAIVSGSALAGVYDQVTRVVTHEMAHAGSPDSTLSIYOGSANAALALQ 60
 DB 1 MKLKVAAFAAIVSGSALAGVYDQVTRVVTHEMAHAGSPDSTLSIYOGSANAALALQ 60
 QY 61 SDAKSEETITGSGYNGADYVGQADNSTIELTONGFRNNATIDQNNAKNSDITVGOYGG 120
 DB 61 SDAKSEETITGSGYNGADYVGQADNSTIELTONGFRNNATIDQNNAKNSDITVGOYGG 120
 QY 121 NNAALVNOTASDSSVMVROVGFNNATANQY 151
 DB 121 NNAALVNOTASDSSVMVROVGFNNATANQY 151

RESULT 2

AAR74625 AAR74625 standard; protein; 151 AA.

AC AAR74625;
 DT 25-MAR-2003 (revised)
 DT 26-JUN-1995 (first entry)
 DE Agfa sequence.
 KM *Salmonella*; Agfa; vaccine.
 KM *Salmonella*;
 OS *Salmonella*.
 PN WO9425598-A2.
 PD 10-NOV-1994.
 PF 26-APR-1994; 94MO-IB000207.
 PR 26-APR-1993; 93US-00054452.

XX (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
 PA (KING/) KING J.
 PI Key WW, Collinson SK, Clouthier SC, Doran JL,
 DR WPI; 1994-358275/44.
 DR N-PSDB; AAC87467.

PT Eliciting an immune response to *Salmonella* - using attenuated *Salmonella*
 PT strains, vector constructs, or compsns. contg. fimbrial type proteins.
 XX Disclosure; Fig 7B; 95pp; English.
 XX The *Salmonella* Agfa protein and DNA are used in vaccine and genetic

CC immunization compositions, respectively, to elicit an immune response to
 CC *Salmonella* in animals (e.g. food producing animals) and humans. (Updated
 CC on 25-MAR-2003 to correct PN field.)

Sequence 151 AA;

Query Match 86.9%; Score 658; DB 2; Length 151;
 Best Local Similarity 90.1%; Pred. No. 1.6e-57;
 Matches 136; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

QY 1 MKLKVAAFAAIVSGSALAGVYDQVTRVVTHEMAHAGSPDSTLSIYOGSANAALALQ 60
 DB 1 MKLKVAAFAAIVSGSALAGVYDQVTRVVTHEMAHAGSPDSTLSIYOGSANAALALQ 60
 QY 61 SDAKSEETITGSGYNGADYVGQADNSTIELTONGFRNNATIDQNNAKNSDITVGOYGG 120
 DB 61 SDAKSEETITGSGYNGADYVGQADNSTIELTONGFRNNATIDQNNAKNSDITVGOYGG 120
 QY 121 NNAALVNOTASDSSVMVROVGFNNATANQY 151
 DB 121 NNAALVNOTASDSSVMVROVGFNNATANQY 151

RESULT 3

AAB36341 AAB36341 standard; protein; 151 AA.

AC AAB36341;
 DT 26-FEB-2001 (first entry)
 DE *Salmonella enteritidis* Agfa amino acid sequence SEQ ID NO:5.
 KM *Salmonella*; agfa; chromosomal gene replacement; fimbria; epitope;
 KM vaccine; immune response; immunogen.
 OS *Salmonella enteritidis*.
 PN WO200060102-A2.
 PD 12-OCT-2000.

PF 05-APR-2000; 2000WO-CA000356.
 PR 05-APR-1999; 99US-0127888P.

PA (UYVI-) UNIV VICTORIA.

PI White AP, Doran JL, Collinson SK, Kay WW;
 DR WPI; 2000-672631/65.
 DR N-PSDB; AAC64617.

PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.

PS Disclosure; Page 135; 139pp; English.

XX The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAFF) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or

CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fibrillar presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fibrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fibrillar subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fibrillae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention

SQ Sequence 151 AA;

Query Match 86.9%; Score 658; DB 3; Length 151;
Best Local Similarity 90.1%; Pred. No. 1.6e-57;
Matches 136; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

QY 1 MLLKVAAPAAIVSGSALAGVYDQVTRVTHMAHAGSPDSTSIYOGSANAALALQ 60
DB 1 MLLKVAAPAAIVSGSALAGVYDQVTRVTHMAHAGSPDSTSIYOGSANAALALQ 60
QY 61 SPARKSETTITGSGYNGADVQGADNSTIELTONGFRNNATTIDQNAKNSDITVGOYGG 120
DB 61 SPARKSETTITGSGYNGADVQGADNSTIELTONGFRNNATTIDQNAKNSDITVGOYGG 120
QY 121 NNAALVNOTASDSSVMVROVGFNNATANQY 151
DB 121 NNAALVNOTASDSSVMVROVGFNNATANQY 151

RESULT 4

ID AAM23570 standard; protein; 151 AA.

AC AAM23570;

DT 25-MAR-2003 (revised)

DT 29-SEP-1997 (first entry)

DE Salmonella enteritidis 27655-3b agfa.

KW Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody.

OS Salmonella enteritidis.

FH Key Location/Qualifiers

FT Misc-difference 123

FT /note= "Encoded by GCC"

PN US5635617-A.

PD 03-JUN-1997.

PF 26-APR-1994; 94US-00233788.

PR 26-APR-1993; 93US-00054452.

PA (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.

PI Collinson SK, Kay WM, Doran JL;

DR MPI: 1997-309886/28.

DR N-PSDB; AAT74142.

XX Isolated Salmonella gene agfa - used for diagnosis of Salmonella or

XX Enteropathogenic bacteria of the Enterobacteria family.

XX Example 2; Fig 7; 85pp; English.

XX The present sequence represents agfa encoded by the full agfa gene
CC derived from Salmonella enteritidis 27655-3b. The nucleic acid can be

CC used to provide diagnostic assays for Salmonella and/or enteropathogenic
CC bacteria of the family Enterobacteria. It can also be used to provide
CC proteins and antibodies which can be used for assays. The nucleic acid
CC sequence can be used to provide probes or primers which can specifically
CC hybridise to nucleic acid molecules from greater than 99% of Salmonella
CC strains that are pathogenic to warm-blooded animals relative to nucleic
CC acid molecules from virtually all other microbial organisms. (Updated on
CC 25-MAR-2003 to correct PF field.)

SQ Sequence 151 AA;

Query Match 86.3%; Score 653; DB 2; Length 151;
Best Local Similarity 89.4%; Pred. No. 5.1e-57;
Matches 135; Conservative 3; Mismatches 13; Indels 0; Gaps 0;

QY 1 MLLKVAAPAAIVSGSALAGVYDQVTRVTHMAHAGSPDSTSIYOGSANAALALQ 60
DB 1 MLLKVAAPAAIVSGSALAGVYDQVTRVTHMAHAGSPDSTSIYOGSANAALALQ 60
QY 61 SPARKSETTITGSGYNGADVQGADNSTIELTONGFRNNATTIDQNAKNSDITVGOYGG 120
DB 61 SPARKSETTITGSGYNGADVQGADNSTIELTONGFRNNATTIDQNAKNSDITVGOYGG 120
QY 121 NNAALVNOTASDSSVMVROVGFNNATANQY 151
DB 121 NNAALVNOTASDSSVMVROVGFNNATANQY 151

RESULT 5

ID AAB36349 standard; protein; 151 AA.

AC AAB36349;

DT 26-FEB-2001 (first entry)

DE Agfa: PT3#4 amino acid sequence SEQ ID NO:18.

KW Salmonella; agfa; chromosomal gene replacement; fibrin; epitope;

KW vaccine; immune response; immunogen.

OS Salmonella enteritidis.

OS Escherichia coli.

OS Synthetic.

PN WO200060102-A2.

PD 12-OCT-2000.

PF 05-APR-2000; 2000WO-CA000356.

PR 05-APR-1999; 99US-0127888P.

PA (UYVI-) UNIV VICTORIA.

PI White AP, Doran JL, Collinson SK, Kay WM;

DR MPI: 2000-672631/65.

DR N-PSDB; AAC64625.

XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
XX which encodes foreign epitope or antigen, expresses recombinant Agfa
XX protein useful for eliciting immune response in animal.

XX Disclosure; Page 136; 139pp; English.

XX The present invention describes a recombinant agfa gene (1) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fibrillae (SEPT17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fibrillae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fibrin subunits, respectively; (2)

directing recombination of a recombinant gene into the chromosome of the homologous species; (3) directing recombination of a recombinant gene back into the chromosome of the homologous species, replacing the native copy of that gene; and (4) eliciting an immune response in an animal, comprising separating an amino acid polymer comprising a recombinant Agfa protein containing a replacement segment or segments of foreign amino acid sequence or sequences grown on a *Salmonella*, *E. coli* or *Enterobacteriaceae* host cell, from the host cell and introducing the polymer into the animal in conjunction with a carrier or diluent. (1) is useful for the expression of recombinant Agfa protein which is useful for eliciting an immune response in an animal. In a fibrillar presentation system the heterologous antigens are presented in high numbers (up to 500,000 copies/cell), the hybrid fibrin protein possesses both the immunogenicity and adhesion properties relevant for an efficient live vaccine, the carrier fibrillar subunit proteins are usually strong immunogens, which may be important for directing an immune response against the inserted epitope, and hybrid fibrillae are easy and inexpensive to purify in large amount. The present sequence is given in the exemplification of the present invention

Sequence 151 AA;

Query Match 81.6%; Score 618; DB 3; Length 151;
Best Local Similarity 76.6%; Pred. No. 1.6e-53;
Matches 131; Conservative 0; Mismatches 0; Indels 40; Gaps 2;

QY 1 MRLKVAAPFAIVSGSALAGV-----YDOLVTRVVTHEMAHAGS 40
DB 1 MRLKVAAPFAIVSGSALAGVPMWGGGNGHNGGSSGPDYDOLVTRVVTHEMAH--- 57
QY 41 PDSTLSTIYOGSANAALALQSDARKSETTITQSGYGADVGQADNSTIELTONGFRNN 100
DB 58 -----ALQSDARKSETTITQSGYGADVGQADNSTIELTONGFRNN 100
QY 101 ATTDQNAKNSDITVGYGNGNAALVNOTASDSSVMRQVFGNNATNAY 151
DB 101 ATTDQNAKNSDITVGYGNGNAALVNOTASDSSVMRQVFGNNATNAY 151

RESULT 6
AAB36346
ID AAB36346 standard; protein; 151 AA.

AC AAB36346;

DT 26-FEB-2001 (first entry)

DE Agfa::PT3#1 amino acid sequence SEQ ID NO:12.

KW *Salmonella*; agfa; chromosomal gene replacement; fibrin; epitope;
KM vaccine; immune response; immunogen.

OS *Salmonella enteritidis*.

OS *Escherichia coli*.

OS Synthetic.

PN WO200060102-A2.

PD 12-OCT-2000.

PF 05-APR-2000; 2000WO-CA000356.

PR 05-APR-1999; 99US-0127888P.

PA (UYVI-) UNIV VICTORIA.

PI White AP, Doran JL, Collison SK, Kay MW,

XX WPI; 2000-672631/65.

XX N-PSDB; AAC64622.

Recombinant agfa gene having a segment replaced by a foreign DNA sequence which encodes foreign epitope or antigen, expresses recombinant Agfa

protein useful for eliciting immune response in animal.
Disclosure; Page 135; 139p; English.

The present invention describes a recombinant agfa gene (1) where a segment of the gene has been replaced by a segment of a foreign DNA sequence which encodes a foreign epitope or antigen. Also described are: (1) use of thin aggregative fibrillae (SEF17/TAf) nucleation depended assembly system of strains of *Salmonella*, *Escherichia coli* and *Enterobacteriaceae* for the production of fibrillae comprising recombinant Agfa, CsgA and Agfa-homologue fibrin subunits, respectively; (2) directing recombination of a recombinant gene into the chromosome of the homologous species; (3) directing recombination of a recombinant gene back into the chromosome of the homologous species, replacing the native copy of that gene; and (4) eliciting an immune response in an animal, comprising separating an amino acid polymer comprising a recombinant Agfa protein containing a replacement segment or segments of foreign amino acid sequence or sequences grown on a *Salmonella*, *E. coli* or *Enterobacteriaceae* host cell, from the host cell and introducing the polymer into the animal in conjunction with a carrier or diluent. (1) is useful for the expression of recombinant Agfa protein which is useful for eliciting an immune response in an animal. In a fibrillar presentation system the heterologous antigens are presented in high numbers (up to 500,000 copies/cell), the hybrid fibrin protein possesses both the immunogenicity and adhesion properties relevant for an efficient live vaccine, the carrier fibrillar subunit proteins are usually strong immunogens, which may be important for directing an immune response against the inserted epitope, and hybrid fibrillae are easy and inexpensive to purify in large amount. The present sequence is given in the exemplification of the present invention

Sequence 151 AA;

Query Match 76.6%; Score 580; DB 3; Length 151;
Best Local Similarity 80.1%; Pred. No. 9.8e-50;
Matches 121; Conservative 7; Mismatches 23; Indels 0; Gaps 0;

QY 1 MRLKVAAPFAIVSGSALAGVYDOLVTRVVTHEMAHSGPDSLTSTIYOGSANAALAQ 60
DB 1 MRLKVAAPFAIVSGSALAGVPMWGGGNGHNGGSSGPDSTLSTIYOGSANAALAQ 60
QY 61 SDARKSETTITQSGYGADVGQADNSTIELTONGFRNNATIDQNAKNSDITVGYG 120
DB 61 SDARKSETTITQSGYGADVGQADNSTIELTONGFRNNATIDQNAKNSDITVGYG 120
QY 121 NNAALVNOTASDSSVMRQVFGNNATNAY 151
DB 121 NNAALVNOTASDSSVMRQVFGNNATNAY 151

RESULT 7
AAB36347
ID AAB36347 standard; protein; 151 AA.

AC AAB36347;

DT 26-FEB-2001 (first entry)

DE Agfa::PT3#2 amino acid sequence SEQ ID NO:14.

KW *Salmonella*; agfa; chromosomal gene replacement; fibrin; epitope;
KM vaccine; immune response; immunogen.

OS *Salmonella enteritidis*.

OS *Escherichia coli*.

OS Synthetic.

PN WO200060102-A2.

PD 12-OCT-2000.

PF 05-APR-2000; 2000WO-CA000356.

AAB36352
ID AAB36352 standard; protein; 151 AA.
XX
AC AAB36352;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa::PT#7 amino acid sequence SEQ ID NO:24.
XX
KM Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
XX vaccine; immune response; immunogen.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN MO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI: 2000-672631/65.
DR N-PSDB; AAC64626.
XX
PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
XX protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 138; 139pp; English.
XX
CC The present invention describes a recombinant agfa gene (1) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF1/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SO Sequence 151 AA;
XX
Query Match 76.0%; Score 575; DB 3; Length 151;
Best Local Similarity 81.5%; Pred. No. 3, 1e-49;
Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;
QY 1 MKLLKVAAPAAIVVSGSALAGVVDLVTRVVTHEMAHAGPDSSTISYOGSANAALALQ 60
DB 1 MKLLKVAAPAAIVVSGSALAGVVPWVGCGGNHNGGNSGPDSTISYOGSANAALALQ 60

QY 61 SDARKSETTTGSGYGADVGAGADNSTIELTONGFRNNATTIDKNANKSDITYGQYGG 120
DB 61 SDARKSETTTTGTGCGNGADVGAGADNSTIELTONGFRNNATTIDKNANKSDITLVTRVVT 120
QY 121 NNAALVNOTASDSSVWVROVGFNNATANYQ 151
DB 121 HEMAHANOTASDSSVWVROVGFNNATANYQ 151
RESULT 10
AAB36350
ID AAB36350 standard; protein; 151 AA.
XX
AC AAB36350;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa::PT#5 amino acid sequence SEQ ID NO:20.
XX
KM Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
XX vaccine; immune response; immunogen.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN MO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI: 2000-672631/65.
DR N-PSDB; AAC64626.
XX
PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
XX protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 137; 139pp; English.
XX
CC The present invention describes a recombinant agfa gene (1) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF1/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in

CC the exemplification of the present invention
 XX Sequence 151 AA;
 SQ

Query Match 75.8%; Score 574; DB 3; Length 151;
 Best Local Similarity 80.8%; Pred. No. 3.9e-49;
 Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLKVAAFAIVSGSALAGVYDQLVTRVTHEMAHAGSPDSTLSIYOGSANALALQ 60
 DB 1 MKLKVAAFAIVSGSALAGVYDQLVTRVTHEMAHAGSPDSTLSIYOGSANALALQ 60

QY 61 SDARKSETTITGSGYNGADVCGADNSTIELTONGFRNNATITDQNAKNSDITTVQYGG 120
 DB 61 SDARKSETTITGSGYNGADVCGADNSTIELTONGFRNNATITDQNAKNSDITTVQYGG 120

QY 121 NNAALVNOTASDSVWVROVGFNNATANQY 151
 DB 121 NNAALVNOTASDSVWVROVGFNNATANQY 151

RESULT 11
 AAB36354
 ID AAB36354 standard; protein; 151 AA.
 AC AAB36354;
 DT 26-FEB-2001 (first entry)
 XX
 DE AgfA::PTJ3#9 amino acid sequence SEQ ID NO:28.
 XX
 KW Salmonella; agfA; chromosomal gene replacement; fimbria; epitope;
 KM vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 PN WO200060102-A2.
 PD 12-OCT-2000.
 PF 05-APR-2000; 2000WO-CA000356.
 PR 05-APR-1999; 99US-0127888P.
 PA (UUYI-) UNIV VICTORIA.
 PI White AP, Doran JL, Collison SK, Kay WM;
 DR WPI; 2000-672631/65.
 DR N-PSDB; AAC64630.
 PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant AgfA
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 138; 139pp; English.

CC The present invention describes a recombinant agfA gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/7AP) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC AgfA, Cgga and AgfA-homologue fimbria subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant AgfA
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or

CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

QY Sequence 151 AA;
 SQ

Query Match 75.0%; Score 568; DB 3; Length 151;
 Best Local Similarity 80.1%; Pred. No. 1.5e-48;
 Matches 121; Conservative 4; Mismatches 26; Indels 0; Gaps 0;

QY 1 MKLKVAAFAIVSGSALAGVYDQLVTRVTHEMAHAGSPDSTLSIYOGSANALALQ 60
 DB 1 MKLKVAAFAIVSGSALAGVYDQLVTRVTHEMAHAGSPDSTLSIYOGSANALALQ 60

QY 61 SDARKSETTITGSGYNGADVCGADNSTIELTONGFRNNATITDQNAKNSDITTVQYGG 120
 DB 61 SDARKSETTITGSGYNGADVCGADNSTIELTONGFRNNATITDQNAKNSDITTVQYGG 120

QY 121 NNAALVNOTASDSVWVROVGFNNATANQY 151
 DB 121 NNAALVNOTASDSVWVROVGFNNATANQY 151

RESULT 12
 AAB36351
 ID AAB36351 standard; protein; 151 AA.
 AC AAB36351;
 DT 26-FEB-2001 (first entry)
 XX
 DE AgfA::PTJ3#6 amino acid sequence SEQ ID NO:22.
 XX
 KW Salmonella; agfA; chromosomal gene replacement; fimbria; epitope;
 KM vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 PN WO200060102-A2.
 PD 12-OCT-2000.
 PF 05-APR-2000; 2000WO-CA000356.
 PR 05-APR-1999; 99US-0127888P.
 PA (UUYI-) UNIV VICTORIA.
 PI White AP, Doran JL, Collison SK, Kay WM;
 DR WPI; 2000-672631/65.
 DR N-PSDB; AAC64627.
 PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant AgfA
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 137; 139pp; English.

CC The present invention describes a recombinant agfA gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:

(1) use of thin aggregative fimbriae (SEF17/TAf) nucleation depended assembly system of strains of *Salmonella*, *Escherichia coli* and *Enterobacteriaceae* for the production of fimbriae comprising recombinant Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2) directing recombination of a recombinant gene into the chromosome of the homologous species; (3) directing recombination of a recombinant gene back into the chromosome of the homologous species, replacing the native copy of that gene; and (4) eliciting an immune response in an animal, comprising separating an amino acid polymer comprising a recombinant Agfa protein containing a replacement segment or segments of foreign amino acid sequence or sequences grown on a *Salmonella*, *E. coli* or *Enterobacteriaceae* host cell, from the host cell and introducing the polymer into the animal in conjunction with a carrier or diluent. (1) is useful for the expression of recombinant Agfa protein which is useful for eliciting an immune response in an animal. In a fimbrial presentation system the heterologous antigens are presented in high numbers (up to 500,000 copies/cell), the hybrid fimbrial protein possesses both the immunogenicity and adhesion properties relevant for an efficient live vaccine, the carrier fimbrial subunit proteins are usually strong immunogens, which may be important for directing an immune response against the inserted epitope, and hybrid fimbriae are easy and inexpensive to purify in large amount. The present sequence is given in the exemplification of the present invention

Sequence 151 AA;

Query Match 74.9%; Score 567; DB 3; Length 151;
Best Local Similarity 80.8%; Pred. No. 1.9e-48;
Matches 122; Conservative 5; Mismatches 24; Indels 0; Gaps 0;

QY 1 MKLKVAFPAIVSGSALAGVYDQLVTRVVTHEMAHAGSDSTLSIYQGSANALALQ 60
DB 1 MKLKVAFPAIVSGSALAGVYDQLVTRVVTHEMAHAGSDSTLSIYQGSANALALQ 60
QY 61 SPARKSETTITQSGYGNAGADVGGADNSTIELTONGFRNATTIOMNAXSDITVGGYGG 120
DB 61 SPARKSETTITQSGYGNAGADVGGADNSTIELTONGFRNATTIOMNAXSDITVGGYGG 120
QY 121 NNAAALVNOTASDSSVWVRQVGFNNATANY 151
DB 121 NNAAALVNOTASDSSVWVRQVGFNNATANY 151

RESULT 13

AAB36355
ID AAB36355 standard; protein; 151 AA.

XX AAB36355;

DT 26-FEB-2001 (first entry)

DE Agfa::PT3#10 amino acid sequence SEQ ID NO:30.

XX *Salmonella*; agfa; chromosomal gene replacement; fimbrial; epitope;
KM vaccine; immune response; immunogen.

OS *Salmonella enteritidis*.

OS *Escherichia coli*.

XX Synthetic.

PN WO200060102-A2.

PD 12-OCT-2000.

PF 05-APR-2000; 2000WO-CA000356.

PR 05-APR-1999; 99US-0127888P.

XX (UYVT-) UNIV VICTORIA.

PA White AP, Doran JL, Collison SK, Kay WW;
PI WPI; 2000-672631/65.

DR N-PSDB; AAC64631.

XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.

PS Disclosure; Page 139; 139pp; English.

CC The present invention describes a recombinant agfa gene (1) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAf) nucleation depended
CC assembly system of strains of *Salmonella*, *Escherichia coli* and
CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
CC *Enterobacteriaceae* host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention

Sequence 151 AA;

Query Match 74.8%; Score 566; DB 3; Length 151;
Best Local Similarity 80.8%; Pred. No. 2.4e-48;
Matches 122; Conservative 5; Mismatches 24; Indels 0; Gaps 0;

QY 1 MKLKVAFPAIVSGSALAGVYDQLVTRVVTHEMAHAGSDSTLSIYQGSANALALQ 60
DB 1 MKLKVAFPAIVSGSALAGVYDQLVTRVVTHEMAHAGSDSTLSIYQGSANALALQ 60
QY 61 SPARKSETTITQSGYGNAGADVGGADNSTIELTONGFRNATTIOMNAXSDITVGGYGG 120
DB 61 SPARKSETTITQSGYGNAGADVGGADNSTIELTONGFRNATTIOMNAXSDITVGGYGG 120
QY 121 NNAAALVNOTASDSSVWVRQVGFNNATANY 151
DB 121 NNAAALVNOTASDSSVWVRQVGFNNATANY 151

RESULT 14

AAR62761
ID AAR62761 standard; protein; 120 AA.

XX AAR62761;

DT 27-AUG-2003 (revised)

DT 25-MAR-2003 (revised)

DT 26-JUN-1995 (first entry)

DE Agfa sequence.

XX *Salmonella*; Agfa; vaccine.

OS *Salmonella enteritidis*.

PN WO9425598-A2.
XX 10-NOV-1994.

XX 26-APR-1994; 94WO-IB000207.
XX
XX 26-APR-1993; 93US-00054452.
XX
XX (UUYV-) UNIV VICTORIA INNOVATION & DEV CORP.
XX (KING/) KING J.
XX
XX Kay WW, Collinson SK, Clouthier SC, Doran JL;
XX
XX MPI; 1994-358275/44.
XX N-PSDB; AA073066.
XX
XX Eliciting an immune response to Salmonella - using attenuated Salmonella
XX strains, vector constructs, or compans. contg. fimbrial type proteins.
XX
XX
XX Disclosure; Fig 7A; 95pp; English.
XX
XX The sequence represents the Salmonella enteritidis 27655-3b TnpH α mutant
XX strain Agfa protein. The encoding DNA and isolated Agfa protein are used
XX in genetic immunization and vaccine compositions, respectively, to elicit
XX an immune response to Salmonella in animals (e.g. food producing animals)
XX and humans. (Updated on 25-MAR-2003 to correct PN field.) (Updated on 27-
XX AUG-2003 to correct OS field.)
SQ Sequence 120 AA;

Query Match 64.3%; Score 487; DB 2; Length 120;
Best Local Similarity 97.9%; Pred. No. 1.3e-40; Indels 0; Gaps 0;
Matches 95; Conservative 2; Mismatches 0;

QY 37 HASGPDSTLSIYQGSANALALQSDARKSETTITQSGYGADVGQADNSTIELTONG 96
DB 16 NSSGPDSTLSIYQGSANALALQSDARKSETTITQSGYGADVGQADNSTIELTONG 75
QY 97 FRNNATTIDQWNAKNSDITVGYGNNALVNOTASDS 133
DB 76 FRNNATTIDQWNAKNSDITVGYGNNALVNOTASDS 112

RESULT 15

AAW23569
ID AAW23569 standard; protein; 120 AA.
XX
XX AAW23569;
XX
XX 25-MAR-2003 (revised)
XX 29-SEP-1997 (first entry)
XX
XX Salmonella enteritidis 27655-3b TnpH α mutant agfa fragment.
XX
XX Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody.
XX
XX Salmonella enteritidis.
XX
XX US5635617-A.
XX
XX 03-JUN-1997.
XX
XX 26-APR-1994; 94US-00233788.
XX
XX 26-APR-1993; 93US-00054452.
XX
XX (UUYV-) UNIV VICTORIA INNOVATION & DEV CORP.
XX
XX Collinson SK, Kay WW, Doran JL;
XX
XX MPI; 1997-309886/28.
XX N-PSDB; AAT74141.
XX
XX Isolated Salmonella gene agfa - used for diagnosis of Salmonella or
XX enteropathogenic bacteria of the Enterobacteria family.
XX

PS Example 2, Fig 7, 85pp; English.

XX The present sequence represents an agfa fragment encoded by an agfa gene
XX fragment derived from Salmonella enteritidis 27655-3b TnpH α mutant
XX strain. The nucleic acid can be used to provide diagnostic assays for
XX Salmonella and/or enteropathogenic bacteria of the family Enterobacteria.
XX It can also be used to provide proteins and antibodies which can be used
XX for assays. The nucleic acid sequence can be used to provide probes or
XX primers which can specifically hybridise to nucleic acid molecules from
XX greater than 9% of Salmonella strains that are pathogenic to warm-
XX blooded animals relative to nucleic acid molecules from virtually all
XX other microbial organisms. (Updated on 25-MAR-2003 to correct PF field.)
XX

SQ Sequence 120 AA;

Query Match 64.3%; Score 487; DB 2; Length 120;
Best Local Similarity 97.9%; Pred. No. 1.3e-40; Indels 0; Gaps 0;
Matches 95; Conservative 2; Mismatches 0;

QY 37 HASGPDSTLSIYQGSANALALQSDARKSETTITQSGYGADVGQADNSTIELTONG 96
DB 16 NSSGPDSTLSIYQGSANALALQSDARKSETTITQSGYGADVGQADNSTIELTONG 75
QY 97 FRNNATTIDQWNAKNSDITVGYGNNALVNOTASDS 133
DB 76 FRNNATTIDQWNAKNSDITVGYGNNALVNOTASDS 112

Search completed: March 11, 2004, 18:33:37
Job time : 46.9 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using bw model

Run on: March 11, 2004, 18:25:54 ; Search time 12.4 Seconds
(without alignments)
628.671 Million cell updates/sec

Title: US-09-543-407-16

Perfect score: 757
Sequence: 1 MKLKVAAFAIIVSGSALA.....DSSVVRQVGRGNATANQY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:
1: /cgn2_6/prodata/2/iaa/5A COMB.pep.*
2: /cgn2_6/prodata/2/iaa/5A COMB.pep.*
3: /cgn2_6/prodata/2/iaa/5A COMB.pep.*
4: /cgn2_6/prodata/2/iaa/5A COMB.pep.*
5: /cgn2_6/prodata/2/iaa/5A COMB.pep.*
6: /cgn2_6/prodata/2/iaa/5A COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	653	86.3	151	US-08-233-788A-59	Sequence 59, Appl
2	487	64.3	120	US-08-233-788A-57	Sequence 57, Appl
3	88	11.6	2315	US-09-543-681A-5434	Sequence 5434, Ap
4	86.5	11.4	943	US-09-477-135A-131	Sequence 131, App
5	85	11.2	1415	US-09-252-991A-26438	Sequence 26438, A
6	82	10.8	1026	US-08-194-290-7	Sequence 7, Appl
7	82	10.8	1026	US-08-614-377A-7	Sequence 7, Appl
8	82	10.8	1026	US-09-142-648B-7	Sequence 7, Appl
9	81.5	10.8	746	US-09-134-001C-3214	Sequence 3214, Ap
10	81	10.7	690	US-09-336-115C-22	Sequence 22, Appl
11	81	10.7	691	US-09-336-115C-12	Sequence 12, Appl
12	81	10.7	745	US-09-336-115C-6	Sequence 6, Appl
13	80.5	10.6	1095	US-08-038-682-2	Sequence 69, Appl
14	80.5	10.6	1536	US-08-038-682-2	Sequence 2, Appl
15	80.5	10.6	1536	US-08-302-832-2	Sequence 2, Appl
16	80.5	10.6	1536	US-08-530-198-2	Sequence 2, Appl
17	80.5	10.6	1536	US-08-469-880-2	Sequence 2, Appl
18	80.5	10.6	1536	US-08-728-470-2	Sequence 2, Appl
19	80.5	10.6	1536	US-08-617-697-2	Sequence 2, Appl
20	80.5	10.6	1536	US-08-719-641-2	Sequence 2, Appl
21	80.5	10.6	1536	US-09-206-943-67	Sequence 67, Appl
22	79.5	10.5	720	US-07-731-157A-2	Sequence 2, Appl
23	79.5	10.5	720	US-08-541-780-2	Sequence 2, Appl
24	79.5	10.5	1739	US-09-540-236-3739	Sequence 3739, Ap
25	79	10.4	907	US-07-718-575-2	Sequence 2, Appl
26	79	10.4	907	US-08-481-206-2	Sequence 2, Appl
27	79	10.4	907	US-08-486-269A-2	Sequence 2, Appl

28	78	10.3	1864	2	US-08-804-227C-3	Sequence 3, Appl
29	77.5	10.2	518	3	US-09-043-123-2	Sequence 2, Appl
30	77.5	10.2	2314	4	US-09-268-347-49	Sequence 49, Appl
31	77	10.2	339	4	US-09-252-991A-32096	Sequence 32096, A
32	77	10.2	1612	1	US-08-169-927-2	Sequence 2, Appl
33	77	10.2	10182	4	US-09-134-001C-3159	Sequence 3159, Ap
34	76.5	10.1	892	4	US-09-336-447A-5	Sequence 5, Appl
35	76	10.0	273	4	US-09-328-352-6167	Sequence 6167, Ap
36	76	10.0	741	4	US-09-489-039A-7854	Sequence 7854, Ap
37	76	10.0	906	1	US-08-254-573-2	Sequence 2, Appl
38	76	10.0	906	1	US-08-687-379-2	Sequence 2, Appl
39	76	10.0	906	1	US-08-687-379-4	Sequence 4, Appl
40	76	10.0	906	4	US-08-172-332-1	Sequence 1, Appl
41	76	10.0	906	4	US-08-216-326-2	Sequence 2, Appl
42	76	10.0	943	4	US-09-056-556-204	Sequence 204, App
43	76	10.0	943	4	US-09-072-596-199	Sequence 199, App
44	76	10.0	943	4	US-09-072-967-204	Sequence 204, App
45	75	9.9	359	4	US-09-328-352-6562	Sequence 6562, Ap

ALIGNMENTS

RESULT 1
US-08-233-788A-59
Sequence 59, Application US/08233788A
Patent No. 5635617

GENERAL INFORMATION:
APPLICANT: Doran, James L.
APPLICANT: Kay, William W.
APPLICANT: Collinson, Karen S.
APPLICANT: Clouthier, Sharon C.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
OF SALMONELLA
TITLE OF INVENTION: OF SALMONELLA
NUMBER OF SEQUENCES: 61
CORRESPONDENCE ADDRESS:
ADDRESS: Seed and Berry
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: U.S.A.
ZIP: 98104-7092

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233,788A
FILING DATE: 26-APR-1994
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: King, Joshua
REGISTRATION NUMBER: 35,570
REFERENCE/DOCKET NUMBER: 920043.403C2

TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031

TELEX: 3723836 SEEDANBERY
INFORMATION FOR SEQ ID NO: 59:

SEQUENCE CHARACTERISTICS:
LENGTH: 151 amino acids
TYPE: amino acid
TOPOLOGY: linear

MOLECULE TYPE: protein
US-08-233-788A-59

Query Match 86.3%; Score 653; DB 1; Length 151;
Best Local Similarity 89.4%; Pred. No. 1.8e-60;
Matches 135; Conservative 3; Mismatches 13; Indels 0; Gaps 0;

1 MKLKVAAFAIIVSGSALAGYDQLVTRVVTHEMAHAGSPSTLSIYOGSANAALAIQ 60
|||||

QY AAFALIVVSG-----SALAGYYDQLVTVVTHEMAHSAEFDSTLS-ITYGGSANAA 56
Db ALFVSILVSVNNLIGQNAPAIATEAAYIEQMADVAAMFGYHAGASAAVSALTPEGQALPT 155
QY L-----ALOSDKRSEFTTIQSGLYGGAADVGGAGDNSTIELQNGFRNNATIDDQNAQNSD 112
Db VAGGGALYSAAAQYTTRVFRLGL-AVVGEEN-----VGNGVGNFNIGSANIGNGN 207
QY ITVGYGGNNAALVNQTASDSVMMRVGVGFNNATTAN 149
Db IGSNGISGNIQFGN-VGPGLTALANNIGFNGTSNN 243

US-08-614-377A-7

Query Match 10.8%; Score 82; DB 2; Length 1026;
 Best Local Similarity 24.8%; Pred. No. 7.9;
 Matches 34; Conservative 28; Mismatches 55; Indels 20; Gaps 7;

QY 11 AIVSGSALAGVYDQVTRVTHMAHA---SGPDSLISYOGSANAALALQSDARKSE 67
 DB 463 AAATGATVAGRVAGAVT---ITDSAAASATTAGIATVIGSGAA---TIDSSAL--- 513
 QY 68 TTITSGYGNADVGOGA---DNSTIELTONGFRNNATI-DQWAKNSDITVGOYGN 121
 DB 514 TTVMISGIGSLGIRALATPATNTLTNNVGLTTTGATITDSEAAADGFTTINIAGS 573
 QY 122 NAA--LVNOTASDSSVM 136
 DB 574 TASSTIASLVAAADATTL 590

RESULT 8

US-09-142-648B-7
 Sequence 7, Application US/09142648B
 Patent No. 6210948
 GENERAL INFORMATION:
 APPLICANT: Smit, John
 APPLICANT: Bingle, Made H.
 APPLICANT: No. 6210948elini, John F.
 TITLE OF INVENTION: EXPRESSION AND SECRETION OF HETEROLOGOUS POLYPEPTIDES FROM CAULI
 FILE REFERENCE: 08106/002002
 CURRENT APPLICATION NUMBER: US/09/142,648B
 PRIOR FILING DATE: 1999-03-30
 PRIOR APPLICATION NUMBER: PCT/CA97/00167
 PRIOR FILING DATE: 1997-03-10
 PRIOR APPLICATION NUMBER: US 07/614,377
 NUMBER OF SEQ ID NOS: 12
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 7
 LENGTH: 1026
 TYPE: PRT
 ORGANISM: Caulobacter crescentus
 US-09-142-648B-7

Query Match 10.8%; Score 82; DB 3; Length 1026;
 Best Local Similarity 24.8%; Pred. No. 7.9;
 Matches 34; Conservative 28; Mismatches 55; Indels 20; Gaps 7;

QY 11 AIVSGSALAGVYDQVTRVTHMAHA---SGPDSLISYOGSANAALALQSDARKSE 67
 DB 463 AAATGATVAGRVAGAVT---ITDSAAASATTAGIATVIGSGAA---TIDSSAL--- 513
 QY 68 TTITSGYGNADVGOGA---DNSTIELTONGFRNNATI-DQWAKNSDITVGOYGN 121
 DB 514 TTVMISGIGSLGIRALATPATNTLTNNVGLTTTGATITDSEAAADGFTTINIAGS 573
 QY 122 NAA--LVNOTASDSSVM 136
 DB 574 TASSTIASLVAAADATTL 590

RESULT 9

US-09-134-001C-3214
 Sequence 3214, Application US/09134001C
 Patent No. 6380370
 GENERAL INFORMATION:
 APPLICANT: Lynn Doucette-Stamm et al
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 FILE REFERENCE: GTC-007
 CURRENT APPLICATION NUMBER: US/09/134,001C
 PRIOR FILING DATE: 1998-08-13
 PRIOR APPLICATION NUMBER: US 60/064,964
 PRIOR FILING DATE: 1997-11-08

PRIOR APPLICATION NUMBER: US 60/055,779
 PRIOR FILING DATE: 1997-08-14
 NUMBER OF SEQ ID NOS: 5674
 SEQ ID NO 3214
 LENGTH: 746
 TYPE: PRT
 ORGANISM: Staphylococcus epidermidis
 US-09-134-001C-3214

Query Match 10.8%; Score 81.5; DB 4; Length 746;
 Best Local Similarity 26.5%; Pred. No. 5.7;
 Matches 39; Conservative 12; Mismatches 43; Indels 53; Gaps 7;

QY 7 AFAIIVSGSALAGVYDQVTRVTHE-----MAHSGPDSL-----SI 47
 DB 511 SAFAL-----ANGGTNNNAHSIQVYTHGEGTIEHTSHKMDYTSYMLAETLKGT 565
 QY 48 YQGSANAALALQSDARKSETTITSGYGNADVGOGA-----DNSTIELTON 95
 DB 566 KAYGSAVGH-----GVSVMGAKTGTGTGGYEYKXNLPMNAKDWIN 611
 QY 96 GFNNATIDQWAKNSDITVGOYGN 122
 DB 612 GFSPEYTWVMGPRK--VKQYGTNS 635

RESULT 10

US-09-336-115C-22
 Sequence 22, Application US/09336115C
 Patent No. 6576244
 GENERAL INFORMATION:
 APPLICANT: Wetzlin, Richard A.
 APPLICANT: Guy, Bruno
 TITLE OF INVENTION: LT and CT in Parenteral Immunization
 FILE REFERENCE: 06132/055002
 CURRENT APPLICATION NUMBER: US/09/336,115C
 PRIOR FILING DATE: 1999-06-19
 PRIOR APPLICATION NUMBER: US 09/100,258
 PRIOR FILING DATE: 1998-06-19
 NUMBER OF SEQ ID NOS: 24
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 22
 LENGTH: 690
 TYPE: PRT
 ORGANISM: Helicobacter pylori
 NAME/KEY: SIGNAL
 LOCATION: (1)...(18)
 US-09-336-115C-22

Query Match 10.7%; Score 81; DB 4; Length 690;
 Best Local Similarity 19.0%; Pred. No. 5.7;
 Matches 33; Conservative 27; Mismatches 66; Indels 48; Gaps 5;

QY 8 AFAIIVSGSALAGVYDQVTRV-----VTHEMAHSGPD 42
 DB 99 AYQAVALLNAAVGMQVIALFTGCGPPTNNQSYOSGNTPALNGTTTNCQAYGTSPN 158
 QY 43 STLISYOGSANAALALQSDARKSETTITSGYGNADVGOGADNSTIELTONGFR----- 98
 DB 159 GILSIDIEYKLNQAVOI-----IQTALNONGGMPALNTTGTGCVNIIQOTVYRTTQ 212
 QY 99 NNATIDQWAKNSDITVGOYG-----NNA-LVNOTASDSSVMRQ 139
 DB 213 NNIEHYTENKEIPVSGSSSPFTIQLTYHNAMENLQQAATIMQVILITQ 266

RESULT 11

US-09-336-115C-12
 Sequence 12, Application US/09336115C
 Patent No. 6576244
 GENERAL INFORMATION:


```

; APPLICANT: Weitzin, Richard A.
; APPLICANT: Guy, Bruno
; TITLE OF INVENTION: LT and CT in Parenteral Immunization
; FILE REFERENCE: 06132/055002
; CURRENT APPLICATION NUMBER: US/09/336,115C
; PRIOR FILING DATE: 1999-06-18
; PRIOR APPLICATION NUMBER: US 09/100,258
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PaateSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 691
; TYPE: PR1
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)...(18)
US-09-336-115C-12

```

```

Query Match          10.7%; Score 81; DB 4; Length 691;
Best Local Similarity 19.0%; Pred. No. 5.7;
Matches 33; Conservative 27; Mismatches 66; Indels 48; Gaps 5;

```

```

QY 8 APAIVSGSALAGVTDLVTRV-----VTHEMAHSGPD 42
DB 100 AVOAVLALNAAVGMQVIALFGCGPPTNNQSYSGFNTPALNGTTTCNQAYGTGHN 159
QY 43 STLSIYOGSANAALALQSDARKSETTITOSGNGADVGQGANSTIELTQNGFR---- 98
DB 160 GILSDIDEYKLNQAYQI-----IQALNQGCGMPALNDTTKGVNIIQNTNVTQTQ 213
QY 99 NNATIDQNNAKNSDITVGOYGG-----NNA-LVNGTASDSSVMVWQ 139
DB 214 NNIIHYTENKEKPEIVSYSGGSFPTIQLTYHNAENLQQAATIMQVILITQ 267

```

```

RESULT 12
US-09-336-115C-6
; Sequence 6, Application US/09336115C
; Patent No. 6576244
; GENERAL INFORMATION:
; APPLICANT: Weitzin, Richard A.
; APPLICANT: Guy, Bruno
; TITLE OF INVENTION: LT and CT in Parenteral Immunization
; FILE REFERENCE: 06132/055002
; CURRENT APPLICATION NUMBER: US/09/336,115C
; PRIOR FILING DATE: 1999-06-18
; PRIOR APPLICATION NUMBER: US 09/100,258
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PaateSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 745
; TYPE: PR1
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)...(20)
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 721
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-336-115C-6

```

```

Query Match          10.7%; Score 81; DB 4; Length 745;
Best Local Similarity 23.1%; Pred. No. 6.4;
Matches 27; Conservative 18; Mismatches 68; Indels 4; Gaps 2;
QY 31 VTHEMAHSGPDSTISYOGSANAALALQSDARKSETTITOSGNGADVGQGANSTI 90
DB 152 ITCNPESTGPKMSIDFKLNEAYQIIQQLKNQSGFPELG-GNGTKV---SVNYNY 207

```

```

QY 91 ELTONGFRNATIDQNNAKNSDITVGOYGNNALVNGTASDSSVMVWQVFGNNAT 147
DB 208 ECRQRTADINGVYQPKAKNGSSSSGNGSGSTQRTATTTQDGVITTTNNKAT 264

```

```

RESULT 13
US-09-206-942-69
; Sequence 69, Application US/09206942
; Patent No. 6432669
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High
; FILE REFERENCE: 1038-861 MTS-Jb
; CURRENT APPLICATION NUMBER: US/09/206,942
; PRIOR FILING DATE: 1998-12-08
; PRIOR APPLICATION NUMBER: 09/167,568
; EARLIER FILING DATE: 1998-10-07
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 69
; LENGTH: 1095
; TYPE: PR1
; ORGANISM: Haemophilus influenzae
US-09-206-942-69

```

```

Query Match          10.6%; Score 80.5; DB 4; Length 1095;
Best Local Similarity 26.4%; Pred. No. 12;
Matches 39; Conservative 22; Mismatches 60; Indels 27; Gaps 7;
QY 15 SCSALAGVTDQ-LVTRVVTTH---EMAHSGPDSTISYOGSANAALALQSDARKSET 68
DB 727 TGSILGLESSESSGVTLRTBGAIAVSNISG--NTVTANSGALTTLGSTIKGTESVT 784
QY 69 TITOSGYNGADVGQGANSTIEL-----TONGFRNATIDQNNAKNSDITVGOYGN 121
DB 785 TSSQSG-----DIGTISGTVKATESLTQNSKIKATGGAANVTSATGTIGGTISG 839
QY 122 NAALVNGTASDSSVMVWQVFGNNATYAN 149
DB 840 NTVNTANAGDLTV-----GNGAEIN 860

```

```

RESULT 14
US-08-038-682-2
; Sequence 2, Application US/08038682
; Patent No. 5549897
; GENERAL INFORMATION:
; APPLICANT: BARENKAMP, STEPHEN J
; APPLICANT: ST. GEME IIT, JOSEPH W
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Shoemaker and Maltare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/038,682
; FILING DATE: 16-MAR-1993
; CLASSIFICATION: 424

```

ATTORNEY/AGENT INFORMATION:
NAME: BERSKRESEER, JERRY W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-293
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1536 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-038-682-2

Query Match 10.6%; Score 80.5; DB 1; Length 1536;
Best Local Similarity 26.4%; Pred. No. 20;
Matches 39; Conservative 22; Mismatches 60; Indels 27; Gaps 7;

QY 15 SCSALAGVYDQ-LVTRVYTH---EMAHASGPDSTLSIYOGSANAALALQS-DARKSET 68
DB 1168 TSIILGIGSSSGSVTLTATBEGALAVSNISG--NTVVTANSGALTTLTLAGSTIKGTESVT 1225
QY 69 TITGSGYNGADVGGAGDNSTIEL-----TONGFRNNATIDQNNAKNSDITVGOYGN 121
DB 1226 TSSQSG-----DIGGTISGCTIVEVKATESLTQSNKIKATVGEANVTSATGTIGGTISG 1280
QY 122 NAALVNOTASDSSVMVRQVFGNNATYAN 149
DB 1281 NTVNVTANAGDLTV-----GNGAEN 1301

RESULT 15

US-08-302-832-2
Sequence 2, Application US/08302832
Patent No. 5603938
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Matfare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/302,832
FILING DATE: 16-SEP-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US pct/us93/02166
FILING DATE: 16-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Berskreseer, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-404
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
LENGTH: 1536 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-302-832-2

Query Match 10.6%; Score 80.5; DB 1; Length 1536;
Best Local Similarity 26.4%; Pred. No. 20;
Matches 39; Conservative 22; Mismatches 60; Indels 27; Gaps 7;

QY 15 SCSALAGVYDQ-LVTRVYTH---EMAHASGPDSTLSIYOGSANAALALQS-DARKSET 68
DB 1168 TSIILGIGSSSGSVTLTATBEGALAVSNISG--NTVVTANSGALTTLTLAGSTIKGTESVT 1225
QY 69 TITGSGYNGADVGGAGDNSTIEL-----TONGFRNNATIDQNNAKNSDITVGOYGN 121
DB 1226 TSSQSG-----DIGGTISGCTIVEVKATESLTQSNKIKATVGEANVTSATGTIGGTISG 1280
QY 122 NAALVNOTASDSSVMVRQVFGNNATYAN 149
DB 1281 NTVNVTANAGDLTV-----GNGAEN 1301

Search completed: March 11, 2004, 18:44:50
Job time: 17.4 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:40:56 ; Search time 24.6 Seconds
(without alignments)
1296.103 Million cell updates/sec

Title: US-09-543-407-16

Perfect score: 757
Sequence: 1 MKLKVAAFAAIVSGSALA.....DSSVMROYFGNNTANQY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 809742 seqs, 21153259 residues

Total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: Published Applications_AA.*
2: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	113.5	15.0	445	US-10-369-493-20638	Sequence 20638, A
2	89	11.8	182	US-10-238-075-1549	Sequence 1549, Ap
3	88.5	11.7	597	US-09-793-306-146	Sequence 146, App
4	86.5	11.4	943	US-09-996-634-131	Sequence 131, App
5	86.5	11.4	943	US-09-997-182-131	Sequence 131, App
6	86.5	11.4	943	US-09-997-181-131	Sequence 131, App
7	84	11.1	385	US-09-984-334-1	Sequence 1, Appl
8	84	11.1	385	US-10-004-115A-4	Sequence 34, Appl
9	84	11.1	385	US-10-327-108-1	Sequence 1, Appl
10	83.5	11.0	356	US-10-156-761-9411	Sequence 9411, Ap
11	83.5	11.0	1621	US-10-185-990-10	Sequence 10, Appl
12	83.5	11.0	1626	US-10-185-990-11	Sequence 11, Appl
13	83	11.0	65	US-09-996-194-16	Sequence 16, Appl
14	83	11.0	594	US-10-156-761-13173	Sequence 13173, A
15	82.5	10.9	273	US-10-369-493-20096	Sequence 20096, A

16	82	10.8	534	US-10-369-493-5343	Sequence 5343, Ap
17	82	10.8	1026	US-09-379-931-7	Sequence 7, Appl
18	82	10.8	1026	US-10-223-597-7	Sequence 7, Appl
19	82	10.8	1236	US-09-769-787-109	Sequence 109, App
20	81	10.7	691	US-08-834-666A-12	Sequence 12, Appl
21	81	10.7	691	US-08-834-666A-22	Sequence 22, Appl
22	81	10.7	745	US-08-834-666A-6	Sequence 6, Appl
23	80.5	10.6	1095	US-10-193-764-65	Sequence 65, Appl
24	80.5	10.6	1536	US-10-092-880-2	Sequence 2, Appl
25	80.5	10.6	1536	US-10-193-764-63	Sequence 63, Appl
26	80	10.6	481	US-10-153-668-352	Sequence 352, App
27	80	10.6	503	US-10-153-668-230	Sequence 230, App
28	80	10.6	814	US-10-369-493-3951	Sequence 3951, Ap
29	79.5	10.5	354	US-09-820-843A-21	Sequence 21, Appl
30	79	10.4	552	US-10-369-493-3745	Sequence 3745, Ap
31	78.5	10.4	385	US-10-156-761-13909	Sequence 13909, A
32	78	10.3	278	US-09-810-264-28	Sequence 28, Appl
33	78	10.3	310	US-09-815-242-13976	Sequence 13976, A
34	78	10.3	1376	US-10-238-075-560	Sequence 560, App
35	77.5	10.2	395	US-10-369-493-18788	Sequence 18788, A
36	77.5	10.2	404	US-10-389-566-1029	Sequence 1029, Ap
37	77.5	10.2	518	US-09-976-297-2	Sequence 2, Appl
38	77	10.2	495	US-10-369-493-4429	Sequence 4429, Ap
39	77	10.2	495	US-10-369-493-7187	Sequence 7187, Ap
40	77	10.2	565	US-10-389-566-1717	Sequence 1717, Ap
41	76.5	10.1	388	US-09-738-626-5044	Sequence 5044, Ap
42	76.5	10.1	892	US-09-952-267-5	Sequence 5, Appl
43	76	10.0	210	US-10-369-493-20095	Sequence 20095, A
44	76	10.0	906	US-10-251-661-2	Sequence 2, Appl
45	76	10.0	943	US-10-193-002-199	Sequence 199, App

ALIGNMENTS

RESULT 1
US-10-369-493-20638
; Sequence 20638, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 20638
; LENGTH: 445
; TYPE: PRT
; ORGANISM: Rhodospseudomonas palustris
; FEATURES:
; NAME/KEY: unsure
; LOCATION: (1)..(445)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-20638

Query Match 15.0%; Score 113.5; DB 15; Length 445;

Best Local Similarity 25.0%; Pred.No. 0.00077; Mismatches 63; Indels 57; Gaps 5;

3 LKLVAAFAAIVSGSALAG---VYDOLVTRVTHMHAASGDSLTSLSYOYGSANVALA 58
5 LFFVAVSLVIVSSAAPADSNTVY-----LNGTNDQDQANTTSGNGNSVCA 52
QY 59 -----LQSDARKSETTI-----TQSGNGADVCGGADNSTIET 93

Db 53 FNGNSGFLQENGLTSGANLLTVKQSGNSNVRDIOQKQSGAGNSAIFQEGTGSVELQ 112
 Qy 94 QNGFRRNATIDQNNAX-----SDTTQGYGNNALVNGTASDSSVWV 137
 Db 113 QGTSTGAVPSCGMWMTNDPEVFNKITQDSSSSNGSKSVIQDKNNVFSIKQGTGNTSV 172
 Qy 138 RQVG 141
 Db 173 NQIG 176

RESULT 2

US-10-238-075-1549
 ; Sequence 1549, Application US/10238075
 ; Publication No. US20030148324A1
 ; GENERAL INFORMATION:
 ; APPLICANT: I.N.S.E.R.M.
 ; TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isolated
 ; FILE REFERENCE: BLANDINE
 ; CURRENT APPLICATION NUMBER: US/10/238,075
 ; PRIOR FILING DATE: 2002-09-10
 ; PRIOR APPLICATION NUMBER: 0003145
 ; NUMBER OF SEQ ID NOS: 2000-03-10
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 1549
 ; LENGTH: 182
 ; TYPE: PRT
 ; ORGANISM: *Escherichia coli*
 US-10-238-075-1549

Query Match 11.8%; Score 89; DB 14; Length 182;
 Best Local Similarity 24.5%; Pred. No. 0.11;
 Matches 39; Conservative 29; Mismatches 71; Indels 20; Gaps 7;

Qy 1 MRLKVAFAAIVVSGSALAGVQDLV---RYVTH-MAHAGPDSLTLYOGSANA 56
 Db 3 IKVIALATFVSAVPSGAMA--YDGTITFGKVAQCTVNTSDKDLAVPL---PTVA 56
 Qy 57 LALQDARKSEBT--ITQSGYNGADVGQGA-----DNSTIELTONGFRNNATIDQMA 108
 Db 57 SLMKNMATSGLTFPAIRLTCATGMSAQNVKAYFEPPSSNIDLATNLTAT--PTKA 114
 Qy 109 KNSUITVQGYGNNALVNGTASDSSVWVROVGFNNAT 147
 Db 115 DNVQIQLNSNGSTILLGEADNGQDVQSETIGSDGSAT 153

RESULT 3

US-09-793-306-146
 ; Sequence 146, Application US/09793306
 ; Patent No. US20020098200A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Campos-Neto, Antonio
 ; APPLICANT: Skeiky, Yasir
 ; APPLICANT: Owendale, Pamela
 ; APPLICANT: Jen, Shyan
 ; APPLICANT: Lodes, Michael
 ; APPLICANT: Corixa Corporation
 ; TITLE OF INVENTION: Compounds and Methods for Diagnosis and Immunotherapy
 ; FILE REFERENCE: 014058-008740US
 ; CURRENT APPLICATION NUMBER: US/09/793,306
 ; PRIOR FILING DATE: 2001-02-26
 ; PRIOR APPLICATION NUMBER: US 60/185,037
 ; PRIOR FILING DATE: 2000-02-25
 ; PRIOR APPLICATION NUMBER: US 60/223,828
 ; PRIOR FILING DATE: 2000-08-08
 ; NUMBER OF SEQ ID NOS: 164
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 146
 ; LENGTH: 597

; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: mTTC#3-His
 US-09-793-306-146

Query Match 11.7%; Score 88.5; DB 9; Length 597;
 Best Local Similarity 35.9%; Pred. No. 0.62;
 Matches 28; Conservative 7; Mismatches 32; Indels 11; Gaps 4;

Qy 74 GYGNGADVGQGANSTIELTONGFRNNATIDQ--NMAKSDTTQGYGNNALVNGTAS 131
 Db 387 GFENAGDINGFENAGD--TNTGFGNAGFFNMGIGNAGNEDMGVNGSGSFNVGN--AG 442
 Qy 132 DSSVWVROVGFNNATAN 149
 Db 443 NQS-----VQFGNAGTLN 455

RESULT 4

US-09-996-634-131
 ; Sequence 131, Application US/09996634
 ; Patent No. US20020172684A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Nano, Francis
 ; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
 ; FILE REFERENCE: 61260
 ; CURRENT APPLICATION NUMBER: US/09/996,634
 ; PRIOR FILING DATE: 2001-11-28
 ; PRIOR APPLICATION NUMBER: 09/447,135
 ; PRIOR FILING DATE: 2000-01-03
 ; PRIOR APPLICATION NUMBER: 08/990,823
 ; PRIOR FILING DATE: 1997-12-15
 ; PRIOR APPLICATION NUMBER: US 96/10375
 ; PRIOR FILING DATE: 1996-06-14
 ; PRIOR APPLICATION NUMBER: 60/000,254
 ; PRIOR FILING DATE: 1995-06-15
 ; NUMBER OF SEQ ID NOS: 169
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 131
 ; LENGTH: 943
 ; TYPE: PRT
 ; ORGANISM: *Mycobacterium tuberculosis*
 US-09-996-634-131

Query Match 11.4%; Score 86.5; DB 9; Length 943;
 Best Local Similarity 24.8%; Pred. No. 1.9;
 Matches 39; Conservative 22; Mismatches 73; Indels 23; Gaps 6;

Qy 7 AAFAIVVSG-----SALAGVQDLTVRYVTHMAHAGPDSLTLS-ITYGSAANA 56
 Db 96 ALFVSLVSNLGNAPAIATBAIYQMAQDYAMFGYHAGASAVSALTTPGQALPT 155
 Qy 57 L-----ALQSDARKSETTITQSGYNGADVGQGANSTIELTONGFRNNATIDQMAKNSD 112
 Db 156 VAGGALVSAALAAQVTRVFRNLGL--ANYGEQ-----VGNNGVNFPLSGANIGNGN 207
 Qy 113 ITVGYGNNALVNGTASDSSVWVROVGFNNATAN 149
 Db 208 IGSNITSSNIGFEN-VGPGLTAALNNTIGFNGTGSNN 243

RESULT 5

US-09-997-182-131
 ; Sequence 131, Application US/09997182
 ; Publication No. US20030049263A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Nano, Francis
 ; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
 ; FILE REFERENCE: 61258
 ; CURRENT APPLICATION NUMBER: US/09/997,182

```

; CURRENT FILING DATE: 2001-11-28
; PRIOR APPLICATION NUMBER: 09/447,135
; PRIOR FILING DATE: 2000-01-03
; PRIOR APPLICATION NUMBER: 08/990,823
; PRIOR FILING DATE: 1997-12-15
; PRIOR APPLICATION NUMBER: US 96/10375
; PRIOR FILING DATE: 1996-06-14
; PRIOR APPLICATION NUMBER: 60/000,254
; PRIOR FILING DATE: 1995-06-15
; NUMBER OF SEQ ID NOS: 169
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 131
; LENGTH: 943
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-997-182-131

```

```

Query Match      11.4%; Score 86.5; DB 10; Length 943;
Best Local Similarity 24.8%; Pred. No. 1.9;
Matches 39; Conservative 22; Mismatches 73; Indels 23; Gaps 6;

```

```

Qy 7 AAPAAIVVSG-----SALAGVYDQVTRVYTHEMAHAGSPDSTLS-IYQYGSANNA 56
Db 96 ALFVSLVSNLIGQNPAAIAATEBAVEQWMAQDVAAAMFGTHAGASAASALTPFGQALPT 155
Qy 57 L-----ALQSDARKSETTITQSGYNGADVGGAGDNSTIELTONGFRNNATTIDQWNAKNSD 112
Db 156 VAGGGLVSAALAAQVTRVFRNLGL-ANYGEGN-----VGNQVGNFNLSANIGCN 207
Qy 113 ITVGOYGGNNALVNOQTASDSSVMYRQVFGNNATTAN 149
Db 208 IGSNGIGSSNIGFGN-VGPGLTALNINIGFGNTGSNN 243

```

```

RESULT 6
US-09-997-181-131
; Sequence 131, Application US/09997181
; Publication No. US20030049269A1
; GENERAL INFORMATION:
; APPLICANT: Nano, Francis
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
; FILE REFERENCE: 61257
; CURRENT APPLICATION NUMBER: US/09/997,181
; CURRENT FILING DATE: 2001-11-28
; PRIOR APPLICATION NUMBER: 09/447,135
; PRIOR FILING DATE: 2000-01-03
; PRIOR APPLICATION NUMBER: 08/990,823
; PRIOR FILING DATE: 1997-12-15
; PRIOR APPLICATION NUMBER: US 96/10375
; PRIOR FILING DATE: 1996-06-14
; PRIOR APPLICATION NUMBER: 60/000,254
; PRIOR FILING DATE: 1995-06-15
; NUMBER OF SEQ ID NOS: 169
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 131
; LENGTH: 943
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-997-181-131

```

```

Query Match      11.4%; Score 86.5; DB 10; Length 943;
Best Local Similarity 24.8%; Pred. No. 1.9;
Matches 39; Conservative 22; Mismatches 73; Indels 23; Gaps 6;

```

```

Qy 7 AAPAAIVVSG-----SALAGVYDQVTRVYTHEMAHAGSPDSTLS-IYQYGSANNA 56
Db 96 ALFVSLVSNLIGQNPAAIAATEBAVEQWMAQDVAAAMFGTHAGASAASALTPFGQALPT 155
Qy 57 L-----ALQSDARKSETTITQSGYNGADVGGAGDNSTIELTONGFRNNATTIDQWNAKNSD 112
Db 156 VAGGGLVSAALAAQVTRVFRNLGL-ANYGEGN-----VGNQVGNFNLSANIGCN 207

```

```

Qy 113 ITVGOYGGNNALVNOQTASDSSVMYRQVFGNNATTAN 149
Db 208 IGSNGIGSSNIGFGN-VGPGLTALNINIGFGNTGSNN 243

```

```

RESULT 7
US-09-984-334-1
; Sequence 1, Application US/09984334
; Publication No. US20030105347A1
; GENERAL INFORMATION:
; APPLICANT: WAKITA, RYUHEI
; APPLICANT: ITO, NOBUYA
; TITLE OF INVENTION: METHOD FOR PRODUCING 4-CYANO-3-OXOBUTANOATE AND
; FILE REFERENCE: 7372-72209
; CURRENT APPLICATION NUMBER: US/09/984,334
; CURRENT FILING DATE: 2001-10-29
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 385
; TYPE: PRT
; ORGANISM: Corynebacterium sp.
US-09-984-334-1

```

```

Query Match      11.1%; Score 84; DB 10; Length 385;
Best Local Similarity 25.9%; Pred. No. 1;
Matches 30; Conservative 21; Mismatches 53; Indels 12; Gaps 4;

```

```

Qy 1 MLLKVAAPAAIVVSGSALAGVYDQVTRVYTHEMAHAGSPDSTLSIYQY-GSAPAAAL 59
Db 191 LHLTAAATVIALDVASDKL-----BLATKVGAEV-LSDKQAAENVKRTTSSGALAVL 244
Qy 60 QSDARKSETTITQSGYNGADV-----GGAGDNSTIELTONGFRNNATTIDQWNAK 110
Db 245 DVGQYQPTIDTMAVAGVSDVTYIGIDGQAHAKVFGFSYKASVTPYVGARN 300

```

```

RESULT 8
US-10-004-115A-34
; Sequence 34, Application US/10004115A
; Publication No. US20030134402A1
; GENERAL INFORMATION:
; APPLICANT: ASAKO, HIROYUKI
; APPLICANT: MATSUMURA, KENJI
; APPLICANT: SHIMIZU, MASATOSHI
; APPLICANT: ITO, NOBUYA
; APPLICANT: WAKITA, RYUHEI
; TITLE OF INVENTION: PROCESS FOR PRODUCING OPTICALLY ACTIVE
; FILE REFERENCE: 7372-72249
; CURRENT APPLICATION NUMBER: US/10/004,115A
; CURRENT FILING DATE: 2002-10-23
; PRIOR APPLICATION NUMBER: JP 2000-372704
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: JP 2001-006144
; PRIOR FILING DATE: 2001-01-15
; PRIOR APPLICATION NUMBER: JP 2001-026594
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: JP 2001-175175
; PRIOR FILING DATE: 2001-06-11
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 34
; LENGTH: 385
; TYPE: PRT
; ORGANISM: Corynebacterium sp.
US-10-004-115A-34

```

```

Query Match      11.1%; Score 84; DB 14; Length 385;
Best Local Similarity 25.9%; Pred. No. 1;
Matches 30; Conservative 21; Mismatches 53; Indels 12; Gaps 4;

```

```

Qy 11.1%; Score 84; DB 14; Length 385;
Db 25.9%; Pred. No. 1;
Matches 30; Conservative 21; Mismatches 53; Indels 12; Gaps 4;

```

QY 1 MKLLKVAAPFAIVGSSALAGYVDLVTRVVTTHMAHSGDPSTLSIYQY-GSANAALAL 59
 Db 191 LRHLASAATYIALDVADKL-----ELATKVGAEHV-LSDDAEAENVKRTIGSGGALVL 244
 QY 60 QSDARKSETTITQSGYGADY-----GQADNSTIELTONGFNNATTIQMAKN 110
 Db 245 DFVGYOPTIDTAMAAVAGVSDVTIVIGIDGGAHAKVGFOSPVEASVTPYWGARN 300

RESULT 9
 US-10-327-108-1
 ; Sequence 1, Application US/10327108
 ; Publication No. US20030186400A1
 ; GENERAL INFORMATION:
 ; APPLICANT: ASAKO, HIROYUKI
 ; APPLICANT: MAKITA, RYUHEI
 ; APPLICANT: ITOH, NOBUYA
 ; TITLE OF INVENTION: METHOD FOR PRODUCING OPTICALLY ACTIVE
 ; FILE REFERENCE: 2-HYDROXYCYCLOALKANECARBOXYLIC ACID ESTER
 ; CURRENT APPLICATION NUMBER: US/10/327,108
 ; PRIOR FILING DATE: 2002-12-24
 ; PRIOR APPLICATION NUMBER: JP 2001-395884
 ; PRIOR FILING DATE: 2001-12-27
 ; PRIOR APPLICATION NUMBER: JP 2001-395885
 ; PRIOR FILING DATE: 2001-12-27
 ; PRIOR APPLICATION NUMBER: JP 2002-107648
 ; NUMBER OF SEQ ID NOS: 15
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 1
 ; LENGTH: 385
 ; TYPE: PRT
 ; ORGANISM: Corynebacterium pseudodiphtheriticum
 US-10-327-108-1

Query Match 11.1%; Score 84; DB 14; Length 385;
 Best Local Similarity 25.9%; Pred. No. 1;
 Matches 30; Conservative 21; Mismatches 53; Indels 12; Gaps 4;
 QY 1 MKLLKVAAPFAIVGSSALAGYVDLVTRVVTTHMAHSGDPSTLSIYQY-GSANAALAL 59
 Db 191 LRHLASAATYIALDVADKL-----ELATKVGAEHV-LSDDAEAENVKRTIGSGGALVL 244
 QY 60 QSDARKSETTITQSGYGADY-----GQADNSTIELTONGFNNATTIQMAKN 110
 Db 245 DFVGYOPTIDTAMAAVAGVSDVTIVIGIDGGAHAKVGFOSPVEASVTPYWGARN 300

RESULT 10
 US-10-156-761-9411
 ; Sequence 9411, Application US/10156761
 ; Publication No. US20030119018A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OMURA, SATOSHI
 ; APPLICANT: IKEDA, HARUO
 ; APPLICANT: ISHIKAWA, JUN
 ; APPLICANT: HORIKAWA, HIROSHI
 ; APPLICANT: SHIBA, TADATOSHI
 ; APPLICANT: SAKAKI, YOSHIYUKI
 ; APPLICANT: HATORI, MASAHIRA
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-262
 ; CURRENT APPLICATION NUMBER: US/10/156,761
 ; PRIOR FILING DATE: 2002-05-29
 ; PRIOR APPLICATION NUMBER: JP 2001-204089
 ; PRIOR FILING DATE: 2001-05-30
 ; PRIOR APPLICATION NUMBER: JP 2001-272697
 ; NUMBER OF SEQ ID NOS: 15109
 ; SEQ ID NO 9411
 ; LENGTH: 356
 ; TYPE: PRT

ORGANISM: Streptomyces avermectilis
 US-10-156-761-9411

Query Match 11.0%; Score 83.5; DB 14; Length 356;
 Best Local Similarity 28.0%; Pred. No. 1.1;
 Matches 44; Conservative 14; Mismatches 58; Indels 41; Gaps 7;
 QY 11 AIVVGSALAGYVDLVTRVVTTHMAHSGDPSTLSIYQYGSANAALALQSDA 63
 Db 5 ARVLSATYIALGALGVAAPASADPATEVRPRSVAPCGAVTISV-----SCDRTGERSDA 60
 QY 64 -----RKSETTITQSGYGADYGCADNSTIELTONG-FNNAT-----ID--- 104
 Db 61 IEAGPQSEETTERQLRGNDNDGAADGETARMPDSDPFGCATGAAGPPAQWVDQC 120
 QY 105 -----QNNKNSDITVGYGGNNAALVNOTASD 132
 Db 121 PVTYGGQESQWTALPSDTRDGE--ARNGAADPSGAD 155

RESULT 11
 US-10-185-990-10
 ; Sequence 10, Application US/10185990
 ; Publication No. US20030073109A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Pan, Jae-Gu
 ; TITLE OF INVENTION: Pan, Jae-Gu ET AL
 ; FILE REFERENCE: 02589, 000100
 ; CURRENT APPLICATION NUMBER: US/10/185,990
 ; CURRENT FILING DATE: 2002-06-28
 ; NUMBER OF SEQ ID NOS: 11
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 10
 ; LENGTH: 1621
 ; TYPE: PRT
 ; ORGANISM: BACILLUS SUBTILIS
 US-10-185-990-10

Query Match 11.0%; Score 83.5; DB 14; Length 1621;
 Best Local Similarity 26.7%; Pred. No. 8.6;
 Matches 44; Conservative 21; Mismatches 51; Indels 49; Gaps 11;
 QY 6 VAAFAIVVSG--SALAGYVDLVTR-----VTHMAHSGDPSTLSIYQYGSANAAL 57
 Db 411 IAGYGSTQSGSDSALTAGYGSTQTAQESNLTAGYGSTGTAGADSSL-IAGYGSTQTS- 468
 QY 58 ALQSDARKSETTITQSGYGN-----GADYGCADNSTIELTONGFNNATI 103
 Db 469 -----GSESSLT-AGYGSTOTAREGSTLTAGYGSTGTAGADSSL-----GYGSTOT- 515
 QY 104 DOWNAKNSDITVGYGGNNAALVNOTASDSSVMARQVFGNNATA 148
 Db 516 -----SGSESSLTNG-YGS-----TOTPAQGSVLTIS--GYGSTOTFA 548

RESULT 12
 US-10-185-990-11
 ; Sequence 11, Application US/10185990
 ; Publication No. US20030073109A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Pan, Jae-Gu
 ; TITLE OF INVENTION: JAE GU PAN ET AL
 ; FILE REFERENCE: 02589, 000100
 ; CURRENT APPLICATION NUMBER: US/10/185,990
 ; CURRENT FILING DATE: 2002-06-28
 ; NUMBER OF SEQ ID NOS: 11
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 11
 ; LENGTH: 1626
 ; TYPE: PRT
 ; ORGANISM: Bacillus subtilis
 US-10-185-990-11

THIS PAGE BLANK (USP)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:24:14 ; Search time 10.3 Seconds

(without alignments)
1410.186 Million cell updates/sec

Title: US-09-543-407-16

Perfect score: 757

Sequence: 1 MLLKLVAAFAAIVSGSALA.....DSSVMRVQYFGNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	658	86.9	151	2 J06039	fimbria protein ag
2	658	86.9	151	2 A10635	major curlin chain
3	484	63.9	151	2 S70788	curlin protein csg
4	474.5	62.7	152	2 D90806	curlin major subun
5	474.5	62.7	152	2 H85665	hypothetical prote
6	118.5	15.7	151	2 S70787	curlin nucleator p
7	118.5	15.7	151	2 C90806	minor curlin subun
8	118.5	15.7	151	2 G85665	curlin minor chain
9	111.5	14.7	151	2 J06040	fimbria protein ag
10	111.5	14.7	151	2 A10635	nucleation compone
11	101.5	13.4	590	1 A45621	leishmanolysin (EC
12	101	13.3	599	2 B42049	leishmanolysin (EC
13	101	13.3	599	2 A44951	leishmanolysin (EC
14	99.5	13.1	582	2 F70675	probable PPB prote
15	99.5	13.1	646	1 S19916	leishmanolysin (EC
16	98	12.9	145	2 AD3143	conserved hypothet
17	98	12.9	145	2 H98144	hypothetical prote
18	97.5	12.9	1034	2 J02143	ice nucleation act
19	96.5	12.7	602	1 JC0221	leishmanolysin (EC
20	96	12.7	1651	2 J01340	outer membrane pro
21	95.5	12.6	1258	2 J00188	ice nucleation pro
22	94.5	12.5	1122	2 S07053	ice nucleation pro
23	93.5	12.4	1567	2 S11672	ice nucleation pro
24	91.5	12.1	1555	2 E97835	hypothetical prote
25	89.5	11.8	552	2 D70604	probable PPB prote
26	89.5	11.8	639	2 C42049	leishmanolysin (EC
27	88.5	11.7	329	2 E70946	probable PPB prote
28	87	11.5	329	2 S23247	outer membrane pro
29	87	11.5	331	2 S21406	outer membrane pro

30	87	11.5	331	2 S21408	outer membrane pro
31	87	11.5	455	2 C29349	hypothetical prote
32	87	11.5	573	2 C86266	P3F19.21 protein -
33	86.5	11.4	3300	2 D70575	probable PPB prote
34	85.5	11.3	1210	2 A25547	ice nucleation pro
35	84	11.1	823	2 S14055	nucleoskeletal-lik
36	84	11.1	3716	2 E70969	probable PPB prote
37	83.5	11.0	1200	1 SNPSO	ice nucleation pro
38	82	10.8	528	2 S69589	hypothetical prote
39	82	10.8	534	2 T32020	hypothetical prote
40	82	10.8	1026	2 A48995	paracrystalline su
41	82	10.8	1073	2 C87374	S-layer protein R
42	82	10.8	1635	2 A10452	hemolysin (importe
43	82	10.8	4776	2 B95206	cell wall surface
44	81.5	10.8	760	2 C90739	hypothetical prote
45	81.5	10.8	760	2 B85589	hypothetical prote

ALIGNMENTS

RESULT 1

J06039

fimbria protein agfa precursor - Salmonella enteritidis

C/Species: Salmonella enteritidis

C/Date: 31-Dec-1996 #sequence revision 31-Dec-1996 #ext_change 08-Oct-1999

C/Accession: J06039; PC6015; A44898

R/Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Baner, P.A.; Kay, W.W.

J. Bacteriol. 178, 662-667, 1996

A/Title: Salmonella enteritidis agfA operon encoding thin, aggregative fimbriae.

A/Reference number: J06039; MUID:96146512; PMID:8550497

A/Accession: J06039

A/Molecule type: DNA

A/Residues: 1-151 <COL>

A/Cross-references: GB:U43280; NID:G1184712; PID:AC43599.1; PID:G1184714

A/Accession: PC6015

A/Molecule type: protein

A/Residues: 21-52 <CO2>

A/Experimental source: strain 27655-3b

A/Note: the authors translated the codon ACG for residue 44 as Ile

R.Collinson, S.K.; Emdy, L.; Muller, K.H.; Trust, T.J.; Kay, W.W.

J. Bacteriol. 173, 4773-4781, 1991

A/Title: Purification and characterization of thin, aggregative fimbriae from Salmonella

A/Reference number: A44898; MUID:91310586; PMID:1677357

A/Accession: A44898

A/Contents: 27655

A/Status: preliminary

A/Molecule type: protein

A/Residues: 21-33 <CO3>

A/Note: sequence extracted from NCBI backbone (NCBIP:45936)

C/Genetics:

A/Genes: agfa

C/Punction:

A/Description: major component of thin aggregative fimbriae

A/Note: fimbriae bind to fibronectin, plasminogen, tissue plasminogen activator

C/Keywords: fimbria

F:1-20/Domain: signal sequence #status predicted <SIG>

F:1-151/Product: fimbria protein agfa #status experimental <MAT>

Query Match

Best Local Similarity 86.9%; Score 658; DB 2; Length 151;

Matches 136; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

QY	1	MLLKVAAPFAAIVSGSALAGYDOLVTRVTHMAHAGPSTLSIYOGSANAALATQ	60
DB	1	MLLKVAAPFAAIVSGSALAGYDOLVTRVTHMAHAGPSTLSIYOGSANAALATQ	60
QY	61	SDARKSETTITGSGYNGADVGQADNNTIELTONGFRNATIDQNNAKNSDITVGQYGG	120
DB	61	SDARKSETTITGSGYNGADVGQADNNTIELTONGFRNATIDQNNAKNSDITVGQYGG	120
QY	121	NNAALVNGTASDSSVMRVQYFGNNATANQY 151	
DB	121	NNAALVNGTASDSSVMRVQYFGNNATANQY 151	

Db 121 NNAALVNOTASDSSVMVRQVGFNNATNMQY 151

RESULT 2

AI0635

major curlin chain precursor (imported) - *Salmonella enterica* subsp. *enterica* serovar Typhimurium

C/Species: *Salmonella enterica* subsp. *enterica* serovar Typhimurium

A/Note: this species has also been called *Salmonella typhimurium*

C/Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002

C/Accession: AI0635

R/Parikh, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Conerton, D.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Mout, S.; O'Garra, P.

Nature 413, 848-852, 2001

A/Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.

A/Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhimurium

A/Reference number: AB0502; MUID:21534947; PMID:11677608

C/Accession: AI0635

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-151 <PAR>

A/Cross-References: GB:AL513382; PIDN:CAD08268.1; PID:G16502315; GSPDB:GN00176

C/Genetics:

A/Gene: STY1181

Query Match 86.9%; Score 658; DB 2; Length 151;
Best Local Similarity 90.1%; Pred. No. 7.5e-52;
Matches 136; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

QY 1 MGLKVAAPAAIVVSGSALAGVYDQVTRVVTHEMAHSGPDSITLYOGSANAALALQ 60

Db 1 MGLKVAAPAAIVVSGSALAGVYDQVTRVVTHEMAHSGPDSITLYOGSANAALALQ 60

QY 61 SDARKSETTITQSGYGNADVGQAGDNSTIELTONGFRNNTIOMNAKNSDITVGOYGG 120

Db 61 SDARKSETTITQSGYGNADVGQAGDNSTIELTONGFRNNTIOMNAKNSDITVGOYGG 120

QY 121 NNAALVNOTASDSSVMVRQVGFNNATNMQY 151

Db 121 NNAALVNOTASDSSVMVRQVGFNNATNMQY 151

RESULT 3

S70788

curlin protein csgA precursor - *Escherichia coli* (strain K-12)

N/Alternate names: csgA protein; major curlin protein

C/Species: *Escherichia coli*

C/Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 01-Mar-2002

C/Accession: S70788; G64846; S31202; S34560; S34559

R/Hammar, M.; Arngvist, A.; Bian, Z.; Olsen, A.; Normark, S.

Mol. Microbiol. 18, 661-670, 1995

A/Title: Expression of two csg operons is required for production of fibronectin- and Cc

A/Reference number: S70783; MUID:96414468; PMID:8817489

A/Accession: S70788

A/Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-151 <HAM>

A/Cross-References: EMBL:X90754; NID:G1147558; PIDN:CAA62282.1; PID:G1147564

A/Experimental source: strain K12, substrain W3110

A/Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995

R/Baltner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co

A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A/Title: The complete genome sequence of *Escherichia coli* K-12.

A/Reference number: A64720; MUID:97426617; PMID:9278503

A/Accession: G64846

A/Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-151 <BLAT>

A/Cross-References: GB:AB000205; GB:U00096; NID:G1787265; PIDN:AACT4126.1; PID:G1787279;

A/Experimental source: strain K-12, substrain W6165

R/Olsen, A.; Arngvist, A.; Hammar, M.; Sukupolvi, S.; Normark, S.

Mol. Microbiol. 7, 523-536, 1993

A/Title: The Rpos sigma factor relieves H-NS-mediated transcriptional repression of csg

A/Reference number: S31202; MUID:93211294; PMID:8459772

A/Accession: S31202

A/Molecule type: DNA

A/Residues: 1-6, 'V', 8-151 <OLS>

A/Cross-References: EMBL:U04979

A/Accession: S34560

A/Molecule type: protein

A/Residues: 21-42/44-50 <OLS2>

R/Olsen, A.N.; Arngvist, A.M.

submitted to the EMBL Data Library, October 1992

A/Reference number: S34559

A/Accession: S34559

A/Molecule type: DNA

A/Residues: 1-133, 'RQDSGWLW' <OLS3>

A/Cross-References: EMBL:U04979; NID:G290424; PIDN:AAA23616.1; PID:G290425

A/Experimental source: strain K-12, substrain W3110

C/Genetics:

A/Map position: 23.15

C/Function:

A/Description: major component of wild-type curli; interaction between CsgA and CsgB tr and H-kinogen; in the absence of CsgA, CsgB can self-assemble into polymers

F/1-20/Domain: signal sequence #status: predicted <SIG>

F/21-151/Product: curlin #status: experimental <MAT>

Query Match 63.9%; Score 484; DB 2; Length 151;
Best Local Similarity 66.9%; Pred. No. 2.7e-36;
Matches 101; Conservative 18; Mismatches 32; Indels 0; Gaps 0;

QY 1 MGLKVAAPAAIVVSGSALAGVYDQVTRVVTHEMAHSGPDSITLYOGSANAALALQ 60

Db 1 MGLKVAAPAAIVVSGSALAGVYDQVTRVVTHEMAHSGPDSITLYOGSANAALALQ 60

QY 61 SDARKSETTITQSGYGNADVGQAGDNSTIELTONGFRNNTIOMNAKNSDITVGOYGG 120

Db 61 SDARKSETTITQSGYGNADVGQAGDNSTIELTONGFRNNTIOMNAKNSDITVGOYGG 120

QY 121 NNAALVNOTASDSSVMVRQVGFNNATNMQY 151

Db 121 NNAALVNOTASDSSVMVRQVGFNNATNMQY 151

RESULT 4

D90806

curlin major subunit CsgA [imported] - *Escherichia coli* (strain O157:H7, substrain RIMD

C/Species: *Escherichia coli*

C/Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001

C/Accession: D90806

R/Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G

gasawara, N.; Yasunaga, T.; Kihara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A/Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and gen

A/Reference number: A99629; MUID:21156231; PMID:11258796

A/Accession: D90806

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-152 <HAY>

A/Cross-References: GB:BA000007; PIDN:BA034843.1; PID:G13360880; GSPDB:GN00154

A/Experimental source: strain O157:H7, substrain RIMD 0509952

C/Genetics:

A/Gene: ECs1420

Query Match 62.7%; Score 474.5; DB 2; Length 152;
Best Local Similarity 65.8%; Pred. No. 2e-35;
Matches 100; Conservative 19; Mismatches 32; Indels 1; Gaps 1;

QY 1 MGLKVAAPAAIVVSGSALAGVYDQVTRVVTHEMAHSGPDSITLYOGSANAALALQ 59

Db 1 MGLKVAAPAAIVVSGSALAGVYDQVTRVVTHEMAHSGPDSITLYOGSANAALALQ 60

```
Db      61 QADANRSDLTITQHGGGADVGQSDSDSIDLTQGFNGSATLQDMNKDSHMTVKQFG 120
      120 GNNAAVNVQTASDSSVVRQVGFNNATATNOY 151
      121 GGNAAVNDQTASNSTVNTVTVQVFGNNATATNOY 152

RESULT 5
H85665
hypothetical protein csgA [imported] - Escherichia coli (strain O157:H7, substrain EDL93
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: H85665
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Niller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: AB5480; MUID:21074935; PMID:11206551
A:Accession: H85665
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-152 <STO>
A:Cross-references: GB:AE005174; NID:g12514573; PIDN:AA655788.1; GSPDB:GN00145; UMGF:Z16
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: csgA

Query Match      62.7%; Score 474.5; DB 2; Length 152;
Best Local Similarity 65.8%; Pred. No. 2e-35; Mismatches 32; Indels 1; Gaps 1;
Matches 100; Conservative 19;

Db      1 MGLTVAAFPAAIVGSGALAGVYDQVTRVTVH-EMAHSGPDSLTISYQVGSANALAL 59
      1 MGLTVAAFPAAIVGSGALAGVYDQVTRVTVH-EMAHSGPDSLTISYQVGSANALAL 60
      1 MGLTVAAFPAAIVGSGALAGVYDQVTRVTVH-EMAHSGPDSLTISYQVGSANALAL 60

Qy      60 QSDARKSETTITQSGYGCADVGGQADNSTIELTONGFPNNATITDQNAKNSDITVQYQ 119
      61 QADANRSDLTITQHGGGADVGQSDSDSIDLTQGFNGSATLQDMNKDSHMTVKQFG 120

Db      120 GNNAAVNVQTASDSSVVRQVGFNNATATNOY 151
      121 GGNAAVNDQTASNSTVNTVTVQVFGNNATATNOY 152

RESULT 6
S70787
curlin nucleator protein csgB precursor - Escherichia coli (strain K-12)
N:Alternate names: csgB protein; curlin nucleation component; minor curlin protein
C:Species: Escherichia coli
C:Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 01-Mar-2002
C:Accession: S70787; F64846
R:Hammar, M.; Arngvist, A.; Bian, Z.; Olsen, A.; Normark, S.
Mol. Microbiol. 18, 661-670, 1995
A:Title: Expression of two csg operons is required for production of fibronectin- and Co
A:Reference number: S70783; MUID:96414468; PMID:8817489
A:Accession: S70787
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-151 <HAM>
A:Cross-references: EMBL:X90754; NID:g1147558; PIDN:CAA62281.1; PID:g1147563
A:Experimental source: strain K12, substrain W3110
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: F64846
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-151 <BLAT>
A:Cross-references: GB:AE000205; GB:U00096; NID:g1787265; PIDN:AACT4125.1; PID:g1787278;
```

```
A:Experimental source: strain K-12, substrain M31655
C:Genetics:
A:Gene: csgB
A:Map position: 23.15
C:Function:
A:Description: minor component of wild-type curli; interaction between CsgA and CsgB tri
A:Note: curli are thin, coiled fibers expressed on the surface of Escherichia coli that
and H-kininogen; in the absence of CsgA, CsgB can self-assemble into polymers
F.1-21/Dominant: signal sequence #status predicted <SIG>
F.22-151/Product: minor curlin chain #status predicted <MAT>

Query Match      15.7%; Score 118.5; DB 2; Length 151;
Best Local Similarity 31.6%; Pred. No. 0.0014;
Matches 37; Conservative 13; Mismatches 62; Indels 5; Gaps 2;

Qy      35 MAHAGPDSLTISYQVGSANALALQSDARKSETTITQSGYGCADVGGQADNSTIELTQ 94
      18 IMAAGVYDLANSEYF---AVNELSKSFNQALITGQAGTNNASHLRQGSKLAVVAQ 73

Db      95 NGFRNNATITDQNAKNSDITVQVGSANALALVNVQTASDSSVVRQVGFNNATATNOY 151
      74 EGSSNRAKITDQGYNL-AVTDQASANDASISQAYGNTAMITQKSGNKNATITQY 129

RESULT 7
C90806
minor curlin subunit precursor CsgB [imported] - Escherichia coli (strain O157:H7, subst
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: C90806
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C. G.
Gisawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen
A:Reference number: AB9629; MUID:21156231; PMID:11258796
A:Accession: C90806
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-151 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA834842.1; PID:g13360879; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetics:
A:Gene: EC81419

Query Match      15.7%; Score 118.5; DB 2; Length 151;
Best Local Similarity 31.6%; Pred. No. 0.0014;
Matches 37; Conservative 13; Mismatches 62; Indels 5; Gaps 2;

Qy      35 MAHAGPDSLTISYQVGSANALALQSDARKSETTITQSGYGCADVGGQADNSTIELTQ 94
      18 IMAAGVYDLANSEYF---AVNELSKSFNQALITGQAGTNNASHLRQGSKLAVVAQ 73

Db      95 NGFRNNATITDQNAKNSDITVQVGSANALALVNVQTASDSSVVRQVGFNNATATNOY 151
      74 EGSSNRAKITDQGYNL-AVTDQASANDASISQAYGNTAMITQKSGNKNATITQY 129

RESULT 8
G85665
curlin minor chain precursor, CsgA homolog [imported] - Escherichia coli (strain O157:H7
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: G85665
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Niller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: AB5480; MUID:21074935; PMID:11206551
A:Accession: G85665
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-151 <STO>
A:Cross-references: GB:AE005174; NID:g12514573; PIDN:AA655787.1; GSPDB:GN00145; UMGF:Z16
```

A:Experimental source: strain 0157:H7, substrain EDL933
C:Genetics:
A:Gene: csfB

Query Match 15.7%; Score 118.5; DB 2; Length 151;
Best Local Similarity 31.6%; Pred. No. 0.0014;
Matches 37; Conservative 13; Mismatches 62; Indels 5; Gaps 2;

QY 35 MAHAGPSTLSIYQGSANALALQSPARKSETTITQSGYGNADVQGGADNSTIELTQ 94
DB 18 IAAAGYDLANSSENF---AVNELSKSEFNQAAIIGAGTNSAQRROGSKLLAVYAO 73

QY 95 NGFRNATIDQWNAKSDITVGOYGGNNAALVNOTASDSSVWVRQVFGNNATANOY 151
DB 74 EGSSRAKIDQTDGPNL-AYIDQAGSANDASISQAGANTMTIIOKSGNKANITQY 129

RESULT 9

UC6040
fimbria protein agfB precursor - Salmonella enteritidis
C:Species: Salmonella enteritidis
C:Date: 31-Dec-1996 #sequence_revision 31-Dec-1996 #text_change 08-Oct-1999

C:Accession: J06040
R:Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Baner, P.A.; Kay, W.W.
J. Bacteriol. 178, 662-667, 1996

A:Title: Salmonella enteritidis agfBAC operon encoding thin, aggregative fimbriae.
A:Reference number: J06039; MUID:96146512; PMID:8550497

A:Accession: J06040

A:Molecule type: DNA

A:Residues: 1-151 <COL>

A:Cross-references: GB:U43280; NID:g1184712; PIDN:AAC43598.1; PID:g1184713

A:Experimental source: strain 276755-3b

A:Gene: agfB

C:Genetics:

A:Description: minor component of thin aggregative fimbriae

A>Note: fimbriae bind to fibronectin, plasminogen, tissue plasminogen activator

C:Keywords: fimbria

F:1-21/Domain: signal sequence #status predicted <SIG>

F:22-151/Product: fimbria protein agfB #status predicted <MAT>

Query Match 14.7%; Score 111.5; DB 2; Length 151;
Best Local Similarity 30.5%; Pred. No. 0.0057;
Matches 36; Conservative 17; Mismatches 46; Indels 19; Gaps 5;

QY 51 GSANALALQSPARKSE-----TTTQSGYGNADVQGGADNST-----IELTQ 94
DB 14 GAGIATATNYDLAREYNFAVNELSKSEFNQAAIIGAGTNSAQRROGSKLLAVISQ 73

QY 95 NGFRNATIDQWNAKSDITVGOYGGNNAALVNOTASDSSVWVRQVFGNNATANOY 151
DB 74 EGSSRAKIDQTDGPNL-AYIDQAGSANDASISQAGANTMTIIOKSGNKANITQY 129

RESULT 10
AH0635
nucleation component of curlin monomers [imported] - Salmonella enterica subsp. enterica
C:Species: Salmonella enterica subsp. enterica serovar Typh
A>Note: this species has also been called Salmonella typhi

C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002

C:Accession: AH0635

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connelton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Mout, S.; O'Garra, P.

Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skellon, J.; Stevens, K.; A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar

A:Reference number: AB0502; MUID:21534947; PMID:11677608

A:Accession: AH0635

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-151 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD08267.1; PID:g16502314; GSPDB:GN00176

C:Genetics:
A:Gene: STY1180

Query Match 14.7%; Score 111.5; DB 2; Length 151;
Best Local Similarity 30.5%; Pred. No. 0.0057;
Matches 36; Conservative 17; Mismatches 46; Indels 19; Gaps 5;

QY 51 GSANALALQSPARKSE-----TTTQSGYGNADVQGGADNST-----IELTQ 94
DB 14 GAGIATATNYDLAREYNFAVNELSKSEFNQAAIIGAGTNSAQRROGSKLLAVISQ 73

QY 95 NGFRNATIDQWNAKSDITVGOYGGNNAALVNOTASDSSVWVRQVFGNNATANOY 151
DB 74 EGSSRAKIDQTDGPNL-AYIDQAGSANDASISQAGANTMTIIOKSGNKANITQY 129

RESULT 11

A45621
leishmanolysin (EC 3.4.24.36) precursor - Leishmania donovani
N:Alternate names: surface endopeptidase glycoprotein gp63
C:Species: Leishmania donovani
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 28-Jan-2000

C:Accession: A45621
R:Webb, J.R.; Button, L.L.; McMaster, W.R.
Mol. Biochem. Parasitol. 48, 173-184, 1991

A:Title: Heterogeneity of the genes encoding the major surface glycoprotein of Leishman
A:Reference number: A45621; MUID:92107220; PMID:1762629

A:Accession: A45621

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-590 <WEB>

A:Experimental source: LV9

A>Note: sequence extracted from NCBI backbone (NCBI:74958, NCBI:74959)

C:Function: A:Description: catalyzes the hydrolysis of peptide bonds between two hydrophobic residu

A>Note: the activated form can activate the proenzyme form

C:Superfamily: leishmanolysin

C:Keywords: blocked carboxyl end; cell adhesion; glycoprotein; hydrolase; lipoprotein; i

F:1-39/Domain: signal sequence #status predicted <SIG>

F:40-87/Domain: activation peptide #status predicted <ATP>

F:88-565/Product: leishmanolysin #status predicted <MAT>

F:566-590/Domain: carboxyl-terminal propeptide #status predicted <CTP>

F:48-251,255,321/Binding site: zinc, catalytic (Cys, His, His) (inhibited) #status

F:112-129,178-217,301-373,380-443,393-412,402-477,454-498,503-553,523-546/Dissulfide bon

F:251,255,321/Binding site: zinc, catalytic (His) (active) #status predicted

F:252/Active site: Glu #status predicted

F:287/Binding site: carbonylate (Asn) (covalent) #status predicted

F:565/Modified site: GPI-anchor ethanolamine amidated carboxyl end (Asn) (in mature for

Query Match 13.4%; Score 101.5; DB 1; Length 590;
Best Local Similarity 26.8%; Pred. No. 0.024;
Matches 46; Conservative 20; Mismatches 59; Indels 35; Gaps 9;

QY 8 AFAIVSGSALGAGYDQVLTTRVVTHEMAHAGPSTLSIYQGSANALALQSPARK-- 65
DB 226 AVGVINIPANIASRYDQVLTTRVVTHEMAHAG--FSVYFFDARLIESVRRHDF 281

QY 66 -----SETTTQS--GYGNGA-----DVGGADNSTELQNGFRNATIDQWNAK 109
DB 262 DVPVNSTAVAKAREQYCGTLEYLEMEDQGGAGSASHIKM-----RNAQ--DELMAP 334

QY 110 NSDITVGOYGGNNAALVNOTA--SDSSVWVRQVFGNNA 146
DB 335 ASD--AGYSSALVTAIFODLGFYQADPS--KAEKMPWGNNA 371

RESULT 12
B42049
leishmanolysin (EC 3.4.24.36) precursor, stationary phase - Leishmania chagasi
N:Alternate names: glycoprotein gp63, stationary phase, surface metalloproteinase, strati
C:Species: Leishmania chagasi
C:Date: 10-Mar-1994 #sequence_revision 07-Apr-1994 #text_change 22-Jun-1999

R:Ramamoorthy, R.; Donelson, J.E.; Paetz, K.E.; Maybodi, M.; Roberts, S.C.; Wilson, M.E.
 J. Biol. Chem. 267, 1888-1895, 1992
 A:Title: Three distinct RNAs for the surface protease gp63 are differentially expressed
 A:Reference number: A42049; MUID:92112918; PMID:1370484
 A:Accession: B42049
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-599 <RAM>
 A:Cross-references: GB:M80669; NID:G159324; PIDN:AAA29236.1; PID:G159325
 C:Function:
 A:Note: the activated form can activate the proenzyme form
 C:Superfamily: leishmanolysin
 C:Keywords: cell adhesion; glycoprotein; hydrolase; membrane bound; metalloproteinase; z
 F:1-35/Domain: signal sequence #status predicted <Sig>
 F:40-97/Domain: activation peptide #status predicted <ATP>
 F:98-599/Product: leishmanolysin #status predicted <MTP>
 F:48-261,265,331/Binding site: zinc, catalytic (Cys, His, His, His) (inhibited) #status
 F:122-133,188-227,311-383,390-451,403-422,412-486,463-507,512-562,532-555/Disulfide bond
 F:261,265,331/Binding site: zinc, catalytic (His) #status predicted
 F:262/Active site: Glu #status predicted
 F:297,394/Binding site: carbohydrate (Aen) (covalent) #status predicted

Query Match 13.3%; Score 101; DB 2; Length 599;
 Best Local Similarity 30.2%; Pred. No. 0.27;
 Matches 32; Conservative 12; Mismatches 34; Indels 28; Gaps 3;

QY 8 AFAATVSGSALAGVDDLVTRVTHMMAASGPDSTLSIYGGSANALALQSDAR--- 64
 236 AAGVNIIPANIASRVDLVTRVTHMMAALG-----FVGPFEGKRIEISISNVNH 288

QY 65 -----KSETTITQSGYGAGADV-----GCGADNSTIEL 92
 289 KQFDVPIVNSTAVAKARQYCDLTLEYLEIEDGGAGSAGSHIKM 334

Db 289 KQFDVPIVNSTAVAKARQYCDLTLEYLEIEDGGAGSAGSHIKM 334

RESULT 13
 A4951
 A:leishmanolysin (EC 3.4.24.36) precursor, log phase - leishmania chagasi
 N:Alternate names: glycoprotein gp63; surface metalloproteinase, log phase
 C:Species: leishmania chagasi
 C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 31-Jan-2000
 C:Accession: A44951; A42049
 R:Miller, R.A.; Reed, S.G.; Parsons, M.
 M:J. Biochem. Parasitol. 39, 267-274, 1990
 A:Title: leishmania gp63 molecule implicated in cellular adhesion lacks an Arg-Gly-Asp s
 A:Reference number: A44951; MUID:90205976; PMID:2320059
 A:Accession: A44951
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-599 <ML>
 A:Cross-references: NID:G159322; PIDN:AAA29235.1; PID:G159323
 U:Ramamoorthy, R.; Donelson, J.E.; Paetz, K.E.; Maybodi, M.; Roberts, S.C.; Wilson, M.E.
 J. Biol. Chem. 267, 1888-1895, 1992
 A:Title: Three distinct RNAs for the surface protease gp63 are differentially expressed
 A:Reference number: A42049; MUID:92112918; PMID:1370484
 A:Accession: A42049
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-599 <RAM>
 A:Cross-references: GB:M80672; NID:G159328; PIDN:AAA29238.1; PID:G159329
 A:Note: sequence extracted from NCBI backbone (NCBIN:76040, NCBI:P:76041)
 A:Note: the source is designated as leishmania donovani chagasi
 C:Function:
 A:Description: catalyzes the hydrolysis of peptide bonds between two hydrophobic residue
 A:Note: the activated form can activate the proenzyme form
 C:Superfamily: leishmanolysin
 C:Keywords: blocked carboxyl end; cell adhesion; glycoprotein; hydrolase; lipoprotein; m
 F:1-33/Domain: signal sequence #status predicted <Sig>
 F:40-97/Domain: activation peptide #status predicted <ATP>
 F:98-574/Product: leishmanolysin #status predicted <MTP>
 F:575-599/Domain: carboxyl-terminal propeptide #status predicted <CTP>
 F:48-261,265,331/Binding site: zinc, catalytic (Cys, His, His, His) (inhibited) #status

F:122-139,188-227,311-383,390-451,403-422,412-486,463-507,512-562,532-555/Disulfide bon
 F:261,265,331/Binding site: zinc, catalytic (His) (active) #status predicted
 F:262/Active site: Glu #status predicted
 F:394/Binding site: carbohydrate (Aen) (covalent) #status predicted
 F:574/Modified site: GPI-anchor ethanolamine amidated carboxyl end (Aen) (in mature for

Query Match 13.3%; Score 101; DB 2; Length 599;
 Best Local Similarity 30.2%; Pred. No. 0.27;
 Matches 32; Conservative 12; Mismatches 34; Indels 28; Gaps 3;

QY 8 AFAATVSGSALAGVDDLVTRVTHMMAASGPDSTLSIYGGSANALALQSDAR--- 64
 236 AAGVNIIPANIASRVDLVTRVTHMMAALG-----FVGPFEGKRIEISISNVNH 288

QY 65 -----KSETTITQSGYGAGADV-----GCGADNSTIEL 92
 289 KQFDVPIVNSTAVAKARQYCDLTLEYLEIEDGGAGSAGSHIKM 334

Db 289 KQFDVPIVNSTAVAKARQYCDLTLEYLEIEDGGAGSAGSHIKM 334

RESULT 14
 F70675
 A:probable PPB protein - Mycobacterium tuberculosis (strain H37Rv)
 C:Species: Mycobacterium tuberculosis
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 15-Sep-2003
 C:Accession: F70675
 R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon,
 J.; Connor, R.; Davies, R.; Devlin, K.; Feldwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S
 Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A:Reference number: A70500; MUID:98295987; PMID:9634230
 A:Accession: F70675
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-582 <COL>
 A:Cross-references: GB:Z82099; GB:A1123456; NID:G3261664; PIDN:CAB05045.1; PID:E280737;
 A:Experimental source: strain H37Rv
 C:Genetics:
 A:Gene: PPE

Query Match 13.1%; Score 99.5; DB 2; Length 582;
 Best Local Similarity 23.4%; Pred. No. 0.35;
 Matches 46; Conservative 19; Mismatches 77; Indels 55; Gaps 7;

QY 8 AFAATVSGSALAGVDDLVTRVTHMMAASGPDSTLSIYGGSANALALQSDAR--- 43
 116 AFAATVSGSALAGVDDLVTRVTHMMAASGPDSTLSIYGGSANALALQSDAR--- 175

QY 44 -TLSIYOGSANAALALQSDARKSETTI--TQSGYGN-----GADVQ 83
 176 PVLGIGNGNALGAGGNTGNTLGNITLGSNGNDANLGSNGISNVGGGVGN 235

QY 84 ---GADNSTIEL-----TONGRRNATTDQWAKAKSDITVGYGGNAAALVQTSAD--- 132
 236 GNFSGGNGRAGLPGSGVNGNGLGNSNLGSGNTGNSNVGFNTGNNVGTGAGSGNIGA 295

QY 133 SSVWVRQVFGNNATAN 149
 296 GNTGSSNWGFNGNIGN 312

Db 296 GNTGSSNWGFNGNIGN 312

RESULT 15
 S19916
 A:leishmanolysin (EC 3.4.24.36) precursor - leishmania mexicana
 N:Alternate names: surface metalloproteinase glycoprotein gp63
 C:Species: leishmania mexicana
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 C:Accession: S19916; A48564
 R:Medina-Acosta, E.; Karses, R.E.; Russell, D.
 submitted to the EMBL Data Library, February 1992
 A:Description: Structurally distinct genes for the surface protease (gp63) of leishmani

A/Accession: S19916
 A/Molecule type: mRNA
 A/Residues: 1-646 <MED>
 A/Cross-references: EMBL:X64394; NID:99559; PIDN:CAA4573.1; PID:99560
 R/Medline-Acosta, E.; Kares, R.E.; Russell, D.G.
 Mol. Biochem. Parasitol. 57, 31-45, 1993
 A/Title: Structurally distinct genes for the surface protease of Leishmania mexicana are
 A/Reference number: A48564; MUID:93149206; PMID:8426614
 A/Accession: A48564
 A/Status: preliminary; not compared with conceptual translation
 A/Molecule type: mRNA
 A/Residues: 1-176, 'Q', 178-646 <ME2>
 A/Cross-references: EMBL:X64394; NID:99559
 A/Note: sequence extracted from NCBI backbone (NCBIP:123747)
 C/Genetics:
 A/Gene: gp63-CI
 A/Map position: 700kb chromosomal band
 C/Function:
 A/Description: catalyzes the hydrolysis of peptide bonds between two hydropho
 A/Note: the activated form can activate the proenzyme form
 C/Superfamily: leishmanolysin
 C/Keywords: cell adhesion; glycoprotein; hydrolase; membrane bound; metalloproteinase; z
 F/1-39/Domain: signal sequence #status predicted <SIG>
 F/40-102/Domain: activation peptide #status predicted <ATP>
 F/103-646/Product: leishmanolysin #status predicted <MAT>
 F/604-620/Domain: transmembrane #status predicted <TM1>
 F/48-266,270,336/Binding site: zinc, catalytic (Cys, His, His, His) (inhibited) #status
 F/86-297,399,409,433,445,466,501/Binding site: carbohydrate (Asn) (covalent) #status pre
 F/127-144,193-232,316-388,395-458,409-427,417-492,469-513,518-568,538-561/Disulfide bond
 F/266,270,336/Binding site: zinc, catalytic (His) (active) #status predicted
 F/267/Active site: Glu #status predicted

Query Match 13.1%; Score 99.5; DB 1; Length 646;
 Best Local Similarity 30.8%; Pred. No. 0.4;
 Matches 40; Conservative 15; Mismatches 66; Indels 9; Gaps 4;
 QY 8 AFAIVVSGSALAGVYDQVLTVRVTHMAHAGPDSITSIYOGSANAALALQSDARK-- 65
 Db 241 AVGVINIPANINASRYDQVLTVRVTHMAHAGVSGTF---FGAVGIVQEVPHLRKDF 296
 QY 66 SETTITSGYGNAGDVQGDN-STIFLTONGFRNNA--TIDMNAKNSDITVGYGANN 122
 Db 297 NVSVITSTVYAKAREQYGCNSLEYLIEDQGGAGSAGSHIKKRNADKELMAPASAGYY 356
 QY 123 AALVNGTASP 132
 Db 357 TALTMVFPD 366

Search completed: March 11, 2004, 18:42:08
 Job time : 11.3 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:14:48 ; Search time 6.3 Seconds

(without alignments)
1248.031 Million cell updates/sec

Title: US-09-543-407-16

Perfect score: 757
Sequence: 1 MLLKRVAPAPAIIVSGSALA.....DSVMVRYQFGNNATANY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	658	86.9	151	1 CSQA_SALTY	P55225 salmonella
2	484	63.9	151	1 CSQA_SALTY	P28307 escherichia
3	474.5	62.7	152	1 CSQA_ECOLI	P39424 escherichia
4	118.5	15.7	151	1 CSQA_ECOLI	P39828 escherichia
5	111.5	14.7	151	1 CSQA_SALTY	P55226 salmonella
6	111.5	14.7	151	1 CSQA_SALTY	P55226 salmonella
7	101.5	13.4	590	1 GRP3_LEIDO	P23223 leishmania
8	101	13.3	599	1 GRP3_LEICH	P15706 leishmania
9	99.5	13.1	646	1 GRP3_LEICH	P43150 leishmania
10	99	13.1	1656	1 OMPB_RICJA	O06653 r outer mem
11	97.5	12.9	1034	1 ICEN_PANAN	O47879 pantoea ana
12	96.5	12.7	602	1 ICEN_PANAN	P08148 leishmania
13	95.5	12.6	1258	1 ICEN_ERWHE	P16239 erwania her
14	94.5	12.5	1332	1 ICEN_PANAN	P20469 pantoea ana
15	93.5	12.4	1567	1 ICEN_XANTU	P18127 xanthomonas
16	93	12.3	310	1 HSLD_STRMU	O8CWA3 streptococc
17	91.5	12.1	1655	1 OMPB_RICCN	O9KKA3 r outer mem
18	87	11.5	331	1 OMB3_NEIMB	P30687 neisseria m
19	87	11.5	331	1 OMB3_NEIMB	P30689 neisseria m
20	85.5	11.3	1210	1 ICEN_PSEFL	P09819 pseudomonas
21	84.5	11.2	1196	1 ICEN_PSEFL	O33475 pseudomonas
22	84	11.1	823	1 NSPI_YEAST	P14907 saccharomyc
23	83.5	11.0	1148	1 ICEN_PSEST	O30611 pseudomonas
24	83.5	11.0	1200	1 ICEN_PSEST	P06620 pseudomonas
25	82.5	10.9	504	1 NRPI_BRARE	O90K44 brachydanio
26	82	10.8	955	1 FRU_DROME	O81881 drosophila
27	82	10.8	1025	1 SLAP_CAUCR	P35828 cauliobacter
28	81.5	10.8	422	1 HEAD_BPAPS	O91184 bacteriophag
29	81.5	10.8	760	1 YBIL_ECOLI	P75780 escherichia
30	81.5	10.8	1158	1 R114_HUMAN	P48582 homo sapien
31	81	10.7	856	1 ALER_AZOVI	O9Z5F9 azotobacter
32	80.5	10.6	720	1 G7AC_BREDI	O91506 brevundimon
33	80	10.6	503	1 NRPI_HUMAN	O16556 homo sapien

34	80	10.6	503	1 NRPI_MOUSE	O9W100 mus musculu
35	80	10.6	621	1 GRP3_LEIGU	O00689 leishmania
36	79.5	10.5	311	1 TBAC_PSESP	P07662 pseudomonas
37	79.5	10.5	1300	1 120K_RICRI	P14914 rickettsia
38	79.5	10.5	1654	1 OMBP_RICRI	O53047 r outer mem
39	79	10.4	907	1 GLR1_MOUSE	P23818 mus musculu
40	79	10.4	907	1 GLR1_RAT	P19490 ratcus norv
41	78	10.3	582	1 Y281_MYCPN	P75383 mycoplasma
42	78	10.3	1443	1 OMBP_RICPR	O53020 r outer mem
43	77	10.2	331	1 OMB NEIMB	P30690 neisseria m
44	76.5	10.1	388	1 ARCI_CORGL	O59280 corynebacte
45	76	10.0	303	1 PANE_HALNI	O9H1F0 halobacteri

ALIGNMENTS

RESULT 1
CSQA_SALTY STANDARD; PRT; 151 AA.
ID CSQA_SALTY
AC P55225;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Major curlin subunit precursor (Fimbriin SEF17).
GN CSQA OR AGPA OR STM1144 OR STY1181 OR T11776.
OS Salmonella typhimurium,
OS Salmonella typhi, and
OS Salmonella enteritidis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
NCBI_TaxID=602, 601, 592;
[1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=SR-11;
RX MEDLINE=98117058; PubMed=9457880;
RA Romling U., Bian Z., Hammar M., Sierra W.D., Normark S.;
RT "Curli fibers are highly conserved between Salmonella typhimurium and
RT Escherichia coli with respect to operon structure and regulation."; J.
RL J. Bacteriol. 180:722-731(1998).
[2]
RN SEQUENCE FROM N.A.
RP SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
RC MEDLINE=21534948; PubMed=11677609;
RX McClelland L., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulyaney E.,
RA Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2."; J.
RL Nature 413:852-856(2001).
[3]
RN SEQUENCE FROM N.A.
RP SPECIES=S.typhi; STRAIN=CT18;
RC MEDLINE=21534947; PubMed=11677608;
RX Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebatina M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holtroyd S., Jagels K.,
RA Krogg A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrett B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18."; J.
RL Nature 413:848-852(2001).
[4]
RN SEQUENCE FROM N.A.
RP SPECIES=S.typhi; STRAIN=Ty2 / ATCC 700931;
RC MEDLINE=22531367; PubMed=12644504;
RX Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodoyanni V., Schwartz D.C., Blattner F.R.;

RT "Comparative genomics of *Salmonella enterica* serovar Typhi strains Ty2
RT and CT18.";
RT J. Bacteriol. 185:2330-2337(2003).
RN [5]
RP SEQUENCE FROM N.A.
RX SPECIES=S. enteritidis; STRAIN=27655-3B;
RA MEDLINE=96146512; PubMed=8550497;
RA Collinson S.K., Clouthier S.C., Doran J.L., Baner P.A., Kay W.W.;
RT "Salmonella enteritidis agfBAC operon encoding thin, aggregative
RT fimbriae.";
RL J. Bacteriol. 178:662-667(1996).
RN [6]
RP SEQUENCE OF 21-151 FROM N.A.
RX SPECIES=S. enteritidis; STRAIN=27655-3B;
RA MEDLINE=94013373; PubMed=8104955;
RA Doran J.L., Collinson S.K., Burian J., Santos G., Todd E.C.D.,
RA Munro C.K., Kay C.M., Baner P.A., Peterkin P.I., Kay W.W.;
RT "DNA-based diagnostic tests for *Salmonella* species targeting agfA,
RT the structural gene for thin, aggregative fimbriae.";
RL J. Clin. Microbiol. 31:2263-2273(1993).
RN [7]
RP SEQUENCE OF 21-33.
RX SPECIES=S. enteritidis; STRAIN=27655-3B;
RA MEDLINE=91310586; PubMed=1677357;
RA Collinson S.K., Emeedy L., Mueller K.-M., Trust T.J., Kay W.W.;
RT "Purification and characterization of thin, aggregative fimbriae from
RT *Salmonella enteritidis*.";
RL J. Bacteriol. 173:4773-4781(1991).
CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN.
CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: AJ002301; CA05317.1; -
DR EMBL: AB008749; AL20074.1; -
DR EMBL: AL627269; CAD08268.1; -
DR EMBL: AB016840; AA069399.1; -
DR EMBL: U43280; AAC3599.1; -
DR PIR: JC6039; JC6039.
DR StyGene: SG10608; csGA.
KW Fibria; Signal; Complete proteome.
FT CHAIN 1 20
FT SIGNAL 1 20
FT CONFLICT 134 151
FT FT
SQ SEQUENCE 151 AA; 15305 MM; B7DAC0D16B621359 CRC64;
Query Match 86.9%; Score 658; DB 1; Length 151;
Best Local Similarity 90.1%; Pred. No. 4.9e-50;
Matches 136; Conservative 3; Mismatches 12; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIYVSGSALAGVVDLVTRVYTHMAHAGSPSTLSIYGSANALALQ 60
DB 1 MKLLKVAFAAIYVSGSALAGVVDLVTRVYTHMAHAGSPSTLSIYGSANALALQ 60
QY 61 SDAKSETTTTOSGYNGADVQAGADNSTLELTONGFRNNATTIDOMNKNSTITVGYGG 120
DB 61 SDAKSETTTTOSGYNGADVQAGADNSTLELTONGFRNNATTIDOMNKNSTITVGYGG 120
QY 121 NNAALVNQASDSVWVROVGRNNATANOY 151
DB 121 NNAALVNQASDSVWVROVGRNNATANOY 151
QY 122 NNAALVNQASDSVWVROVGRNNATANOY 151
DB 122 NNAALVNQASDSVWVROVGRNNATANOY 151

RESULT 2
CSGA_ECOLI STANDARD; PRT; 151 AA.
AC P28307;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Major curlin subunit precursor.
GN CSGA OR B1042.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=K12 / W3110.
RA MEDLINE=93211294; PubMed=8459772;
RA Olsen A., Arngvist A.;
RT "The *kps* sigma factor relieves H-NS-mediated transcriptional
RT repression of *csGA*, the subunit gene of fibronectin-binding curli in
RT Escherichia coli.";
RL Mol. Microbiol. 7:523-536(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=K12 / MC4100;
RA MEDLINE=96414468; PubMed=8817489;
RA Hammer M., Arngvist A., Bian Z., Olsen A., Normark S.;
RT "Expression of two *csg* operons is required for production of
RT fibronectin- and Congo red-binding curli polymers in *Escherichia coli*
RT K-12.";
RL Mol. Microbiol. 18:661-670(1995);
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=K12 / MG1655;
RA MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Valdes J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of *Escherichia coli* K-12.";
RL Science 277:1453-1474(1997).
RN [4]
RP SEQUENCE FROM N.A.
RX STRAIN=K12;
RA MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Maeda S., Miki T., Mizuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horikuchi T.;
RT "A 718-kb DNA sequence of the *Escherichia coli* K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA Res. 3:137-155(1996).
RN [5]
RP SEQUENCE OF 21-40.
RX STRAIN=K12 / W3110;
RA MEDLINE=93023873; PubMed=1357528;
RA Arngvist A., Olsen A., Pfeifer J., Russell D.G., Normark S.;
RT "The *Cur* protein activates cryptic genes for curli formation and
RT fibronectin binding in *Escherichia coli* HB101.";
RL Mol. Microbiol. 6:2443-2452(1992).
RN [6]
RP SEQUENCE OF 21-31.
RX MEDLINE=91310586; PubMed=1677357;
RA Collinson S.K., Emeedy L., Trust T.J., Kay W.W.;
RT "Purification and characterization of thin, aggregative fimbriae from
RT *Salmonella enteritidis*.";
RL J. Bacteriol. 173:4773-4781(1991).
CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN.


```
CC CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC CC or send an email to licenses@isb-sib.ch).
CC CC -----
DR DR EMBL; L09479; AAA23616.1; -.
DR DR EMBL; X90754; CAA62282.1; -.
DR DR EMBL; AB000205; AAC74126.1; -.
DR DR EMBL; D90741; BAA35832.1; -.
DR DR EMBL; D90742; BAA35840.1; -.
DR DR PIR; S70788; S70788.
DR DR EcoGene; EG11489; csGA.
FW FW Flmbria; Signal; Complete proteome.
FT FT SIGNAL 1 20
FT FT CHAIN 21 151 MAJOR CURLIN SUBUNIT.
FT FT CONFLICT 7 7 A -> E (IN REF. 1).
SQ SQ SEQUENCE 151 AA; 15049 MW; C003470D208D395F CRC64;
Query Match 63.9%; Score 484; DB 1; Length 151;
Best Local Similarity 66.9%; Pred. No. 4.8e-35;
Matches 101; Conservative 18; Mismatches 32; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIWVSGSLAGVDDLVTVVTTHMAHAGSDSTLSIYGSSANALALQ 60
DB 1 MKLLKVAIAAIIVPVGSLAGVDPVPGYGGGNGHGCGGNNSGNSELNIQYGGNSALALQ 60
QY 61 SDARKSETTTIQSGYGCADGVGGCANDSTIELTONGFPFNNAITIDOMANKSDITYGYGG 120
DB 61 TDAANSDLTIHQHGGGNCADVGGGSDSDISLTQRGFNGSATLDQMGNKSEMTWKOPGG 120
QY 121 NNALLVNQTASDSVMVRQVRFNNATANOY 151
DB 121 NGGAADVDTASNSSVNTQVTFGNNTATAHOY 151
RESULT 3
CSGA_ECO57 STANDARD; PRT; 152 AA.
ID Q9JUZ4;
AC Q9JUZ4;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Major curlin subunit precursor.
GN CSga OR Z1676 OR ECsl420.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxId=83334;
RX [1]
RN SEQUENCE FROM N.A.
RP STRAIN=O157:H7 / ATCC 43895;
RC MEDLINE=21074935; PubMed=11205551;
RA Uhlrich G.A., Keen J.E., Elder R.O.;
RT "Mutations in the csgD promoter associated with variations in curli
RT expression in certain strains of Escherichia coli O157:H7.";
RL Appl. Environ. Microbiol. 67:2367-2370(2001).
[2]
RN SEQUENCE FROM N.A.
RP STRAIN=O157:H7 / EDL933 / ATCC 700927;
RC MEDLINE=21074935; PubMed=11205551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Groetbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RL "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
Nature 409:529-533(2001).
```

```

RP SEQUENCE FROM N.A.
RC STRAIN=0157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M.,
Ra Han C.-G., Ohtsubo E., Nakayama K.,
RA Iida T., Takami H., Honda T.,
RA Kihara S., Shiba T., Hattori M.,
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12."
RL DNA Res. 8:11-22(2001).
CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREPREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN.
CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcement/).license@eb-eib.ch.
CC -----
CC EMBL; AF275733; AKS3212.1; -
CC DR EMBL; AE005315; AAG55788.1; -
CC DR EMBL; AP002554; BAB34843.1; -
CC DR PIR; D90806; D90806.
CC DR PIR; H85665; H85665.
CC KW Findria; Signal; Complete proteome.
CC FT SIGNAL 1 20 BY SIMILARITY
CC FT CHAIN 21 152 MAJOR CURLIN SUBUNIT.
CC SQ SEQUENCE 152 AA; 15099 MW; BE2D94DDE91243 CRC64;

Query March 62.7%; Score 474.5; DB 1; Length 152;
Best Local Similarity 65.8%; Pred. No. 3.2e-34;
Matches 100; Conservative 19; Mismatches 32; Indels 1; Gaps 1;

Oy 1 MKLKVAFAAIVVSGSALAGVYDQLVTRVYTH-EMHASGSDSTLSIYOGSANAALAL 59
Db 1 MKLKVAFAAIAIVSGSALAGVYDVGSGGNGHGGGNGNSPSELNIYOGGNSALAL 60
Oy 60 QSDARKSETTITGSGYGNGADVGGAGDNSTIELTQNGFRNNATIDQWNAKSDITVGGYG 119
Db 61 QADARNSDLRTITGGGNGADVGGSDSSIDLTRGFGNSATLIDQWNGKDSHTVXQFG 120
Oy 120 GNNAAIYNOTRSDSSVWVRQYGFENNAITANDY 151
Db 121 GGNGAAVDQTRASNSTVAVTVQVGFENNAITANDY 152

RESULT 4
CSGB_ECOLI STANDARD; PRT; 151 AA.
AC P39828;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Minor curlin subunit precursor.
CSGB OR B1041 OR Z1675 OR ECS1419.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 83334;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=K12 / MC4100;
RX MEDLINE=96414468; PubMed=8817489;
RA Hammar M., Arnyqvist A., Bian Z., Olsen A., Normark S.;
RT "Expression of two csg operons is required for production of
RT fibronectin- and Congo red-binding curli polymers in Escherichia coli
```

```

DR EMBL; AB000205; AAC74125.1; -.
DR EMBL; D90741; BA35831.1; -.
DR EMBL; AE005315; AAG55787.1; -.
DR EMBL; AP002554; BAB34842.1; -.
DR PIR; C90806; C90806.
DR PIR; G85665; G85665.
DR PIR; S70787; S70787.
DR EcoGene; EG12621; csGB.
KM Fimbrtia; Signal; Complete proteome.
FT SIGNAL 1 21
FT CHAIN 22 151
SQ SEQUENCE 151 AA; 15882 MW; B18D26B964014B8 CRC64;

Query Match 15.7%; Score 118.5; DB 1; Length 151;
Best Local Similarity 31.6%; Pred. No. 0.0015;
Matches 37; Conservative 13; Mismatches 62; Indels 5; Gaps 2

Qy 35 MHASPSDSTLTSTYQYGSANALALQSDARKSTTTTQYGYNGADVGGADNSTIELTQ 94
Db 18 IAAAPAYDIANSEYFN---ANNEISKSSFNQAATIGAGTNNASQLRQGGSKTLAVAQ 73

Qy 95 NGFRNATIDQWNAKSDITVQYGGNNALVNOTPDSVWVRQVFGNNAITANQY 151
Db 74 EGSNRKATIDQYGDVNL-AYIDQAGSANDASISQAGYGTAMTIQKSGNKANITQY 129

RESULT 5
CSGB_SALTI STANDARD; PRT; 151 AA.
ID CSGB_SALTI
AC 0827M3;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Minor curlin subunit precursor.
CSGB OR STY1180 OR T1777.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_Taxid=601;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=21534947; PubMed=11677608;
RX STRAIN=CT18;
RA Parthill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Felwell T., Hamlin N., Haque A., Hen T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrrell B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18."
RL Nature 413:848-852(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=TY2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;
RA Deng W., Lion S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Butland V., Kodyiam V., Schwartz D.C., Blattner F.R.;
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RT and CT18."
RL J. Bacteriol. 185:2330-2337(2003).
CC -I- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
CC CURLIN MONOMERS.
CC -I- SIMILARITY: BELONGS TO THE CSGB/CSGB FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute.

```

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: AL627269; CAD08267.1; -
 KW EMBL: AE016840; AA069400.1; -
 DR Fimbrin; Signal; Complete proteome.
 FT SIGNAL 1 21
 FT CHAIN 22 151 MINOR CURLIN SUBUNIT.
 SQ SEQUENCE 151 AA; 16254 MW; 161C54326E573495 CRC64;

Query Match 14.7%; Score 111.5; DB 1; Length 151;
 Best Local Similarity 30.5%; Pred. No. 0.0058;
 Matches 36; Conservative 17; Mismatches 46; Indels 19; Gaps 5;

QY 51 GSANALALQSDPKRSE-----TTTQSGYGNADVQ-GADNST-----IELNQ 94
 DB 14 GAPGIATATNYDLARSEYNFAVNELSKSFNOAIIQVGTNSARVROGSKLSVISQ 73

QY 95 NGFRNNATIDQWNAKSDIT-VGQYGNNAALVNOTASDSSVWVRQVFGNNATANQY 151
 DB 74 EGGNNRAKVDQ--AGNYNFAVIEQTGNANDASISQAYGNSAIIQKSGNKRANTQY 129

RESULT 6
 CSGB SALTY STANDARD; PRT; 151 AA.

AC P55226;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Minor curlin subunit precursor (Fimbrin SEF17 minor subunit).
 OS CSGB OR AGFB OR STM1143.
 OS Salmonella typhimurium, and
 OS Salmonella enteritidis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 OX NCBI_TaxId=602, 592;
 [1] _TaxId=602, 592;
 [1] SEQUENCE FROM N.A.
 RC SPECIES=S.typhimurium; STRAIN=SR-11;
 RA MEDLINE=98117058; PubMed=9457880;
 RX Romling U., Bian Z., Hamm M., Sieralta W.D., Normark S.;
 RT "Curli fibers are highly conserved between Salmonella typhimurium and
 Escherichia coli with respect to operon structure and regulation.";
 RL J. Bacteriol. 180:722-731(1998).
 [2]
 [2] SEQUENCE FROM N.A.
 RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSCL412 / ATCC 700720;
 RX MEDLINE=21534948; PubMed=11677609;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Lattelle P.,
 Courtney L., Porwollik S., Ali J., Dante M., Du P., Hou S., Layman D.,
 Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
 Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
 Waterston R., Wilson R.K.;
 RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
 LT2.";
 RL Nature 413:852-856(2001).
 [3]
 [3] SEQUENCE FROM N.A.
 RC SPECIES=S.enteritidis; STRAIN=27655-3B;
 RX MEDLINE=96146512; PubMed=8550497;
 RA Collinson S.K., Cloutier S.C., Doran J.L., Baner P.A., Kay W.W.;
 RT "Salmonella enteritidis agfBAC operon encoding thin, aggregative
 fimbriae.";
 RL fimbriae. 178:662-667(1996).
 J. Bacteriol. 178:662-667(1996).
 -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 FIBRINECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
 CURLIN MONOMERS.
 -1- SIMILARITY: BELONGS TO THE CSGB/CSGB FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: AJ002301; CA05316.1; -
 DR EMBL: AE008749; AL20073.1; -
 DR EMBL: U03280; AAC3598.1; -
 DR PIR: J06040; J06040.
 DR StyGene; SG10609; CSGB.
 KW Fimbrin; Signal; Complete proteome.
 FT SIGNAL 1 21
 FT CHAIN 22 151 MINOR CURLIN SUBUNIT.
 SQ SEQUENCE 151 AA; 16182 MW; C0FC5430B8DD361D CRC64;

Query Match 14.7%; Score 111.5; DB 1; Length 151;
 Best Local Similarity 30.5%; Pred. No. 0.0058;
 Matches 36; Conservative 17; Mismatches 46; Indels 19; Gaps 5;

QY 51 GSANALALQSDPKRSE-----TTTQSGYGNADVQ-GADNST-----IELNQ 94
 DB 14 GAPGIATATNYDLARSEYNFAVNELSKSFNOAIIQVGTNSARVROGSKLSVISQ 73

QY 95 NGFRNNATIDQWNAKSDIT-VGQYGNNAALVNOTASDSSVWVRQVFGNNATANQY 151
 DB 74 EGGNNRAKVDQ--AGNYNFAVIEQTGNANDASISQAYGNSAIIQKSGNKRANTQY 129

RESULT 7
 GP63 LEIDO STANDARD; PRT; 590 AA.

AC P23223;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Leishmanolysin precursor (BC 3.4.24.36) (Cell surface protease)
 DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface
 endopeptidase).
 GN GP63.
 OS Leishmania donovani.
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 OX NCBI_TaxId=5561;
 [1] _TaxId=5561;
 [1] SEQUENCE FROM N.A.
 RC STRAIN=LV9;
 RX MEDLINE=92107220; PubMed=1762629;
 RA Webb J.R., Burton L.L., McMaster R.W.;
 RT "Heterogeneity of the genes encoding the major surface glycoprotein
 of Leishmania donovani.";
 RL Mol. Biochem. Parasitol. 48:173-184(1991).
 CC -1- FUNCTION: Has an integral role during the infection of macrophages
 CC in the mammalian host.
 CC -1- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and
 CC P1' and basic residues at P2 and P3'. A model nonpeptide is
 CC cleaved at -Ala-Tyr-|-Ileu-Iys-Lys--.
 CC -1- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
 CC -1- SIMILARITY: Belongs to peptidase family M8.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: M60048; AAA29244.1; -
 DR HSSP: P08148; ILM.

```

DR MEROPS; M08.001;
DR InterPro: IPR006025; Pept M Zn BS.
DR InterPro: IPR001577; Peptidase_M8.
DR Pfam: PF01457; Peptidase_M8; 1.
DR PRINTS; PR00782; LSHMANOLYSIN.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KM Hydroxylase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;
KW Zymogen; Signal; Cell adhesion; GPI-anchor; Lipoprotein.
FT SIGNAL 1 39
FT PROPEP 40 87
FT CHAIN 88 565
FT PROPEP 566 590
FT PROPEP 551 251
FT ACT SITE 252 252
FT METAL 255 255
FT METAL 321 321
FT DISULFID 112 129
FT DISULFID 178 217
FT DISULFID 301 373
FT DISULFID 380 443
FT DISULFID 393 412
FT DISULFID 402 477
FT DISULFID 454 498
FT DISULFID 503 553
FT DISULFID 523 546
FT CARBOHYD 287 287
FT LIPID 565 565
SQ SEQUENCE 590 AA; 62950 MW; 07B315D29659F58 CRC64;

Query Match 13.4%; Score 101.5; DB 1; Length 590;
Best Local Similarity 28.8%; Pred. No. 0.2;
Matches 46; Conservative 20; Mismatches 59; Indels 35; Gaps 9;

QY 8 AFAIVSGSALAGVYDQLVTRVYTHMAHSGPDSLSTIYQGSANAAALQSDAR-- 65
DB 226 AVGVINIPANIASRYDQLVTRVYTHMAHAG--FSVGFEGARILTESISNVRH 281
QY 66 -----SETTITQS--GYNGA-----DVQGANSTLETLTONGFRNNATIDQWNAK 109
DB 282 DVPVINSSTAVAKAREYGCOTLEYLETEDGSGAGSAGSHIKM-----RNAQ-DELMAP 334
QY 110 NSDITVQYGGANNAALVNOTA---SDSSVMRVQYGFENNA 146
DB 335 ASD--AGYYSALTWALFQDLGFYQADFS-KAEEMPWGRNA 371

RESULT 8
GP63 LEICH STANDARD; PRT; 599 AA.
ID GP63 LEICH
AC P15706;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Leishmanolysin precursor (EC 3.4.24.36) (Cell surface protease)
DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface
endopeptidase).
GN GP63.
OS Leishmania chagasi.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=44271;

[1]
SEQUENCE FROM N.A.
MEDLINE=90205976; Pubmed=2320059;
Miller R.A., Reed S.G., Parsons M.;
"Leishmania gp63 molecule implicated in cellular adhesion lacks an
Arg-Gly-Asp sequence.";
Mol. Biochem. Parasitol. 39:267-274(1990).

[2]
SEQUENCE FROM N.A.
MEDLINE=92112918; Pubmed=1370484;
Ramamoorthy R., Donelson J.E., Paetz K.E., Maybodi M., Roberts S.C.,
Wilson M.E.;

```

```

RT "Three distinct RNAs for the surface protease gp63 are differentially
RT expressed during development of Leishmania donovani chagasi
RT promastigotes to an infectious form."
RL J. Biol. Chem. 267:1888-1895(1992).
CC -1- FUNCTION: Has an integral role during the infection of macrophages
CC in the mammalian host.
CC -1- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and
CC P1' and basic residues at P2 and P3'. A model nonpeptide is
CC cleaved at -Ala-Tyr-Lys-Lys-.
CC -1- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -1- SIMILARITY: Belongs to peptidase family M8.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; M80672; AAA29238.1; -.
DR EMBL; M28527; AAA29235.1; -.
DR PIR; A44951; A44951.
DR HSSP; P08148; 1LMF.
DR MEROPS; M08.001; -.
DR InterPro: IPR006025; Pept M Zn BS.
DR InterPro: IPR001577; Peptidase_M8.
DR Pfam; PF01457; Peptidase_M8; 1.
DR PRINTS; PR00782; LSHMANOLYSIN.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KM Hydroxylase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;
KW Zymogen; Signal; Cell adhesion; GPI-anchor; Lipoprotein.
FT SIGNAL 1 39
FT PROPEP 40 97
FT CHAIN 98 574
FT PROPEP 575 599
FT METAL 261 261
FT ACT SITE 262 262
FT METAL 265 265
FT METAL 331 331
FT DISULFID 122 139
FT DISULFID 188 227
FT DISULFID 311 383
FT DISULFID 390 452
FT DISULFID 403 422
FT DISULFID 412 486
FT DISULFID 463 507
FT DISULFID 512 562
FT DISULFID 532 555
FT CARBOHYD 297 297
FT CARBOHYD 394 394
FT LIPID 574 574
SQ SEQUENCE 599 AA; 63848 MW; 746730AE8E2A2E7C CRC64;

Query Match 13.3%; Score 101; DB 1; Length 599;
Best Local Similarity 30.2%; Pred. No. 0.23;
Matches 32; Conservative 12; Mismatches 34; Indels 28; Gaps 3;

QY 8 AFAIVSGSALAGVYDQLVTRVYTHMAHSGPDSLSTIYQGSANAAALQSDAR-- 64
DB 236 AVGVINIPANIASRYDQLVTRVYTHMAHAG--FSVGFEGARILTESISNVRH 288
QY 65 -----KSETTITQSGYNGADV-----GQGANSTLET 92
DB 289 KDFDVPVINSSTAVAKAREYGCOTLEYLETEDGSGAGSAGSHIKM 334

RESULT 9
ID GP63 LEIME STANDARD; PRT; 646 AA.
AC P43150;

```

DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Leishmanolysin C1 precursor (EC 3.4.24.36) (Cell surface protease)
 DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface endopeptidase).
 GN GP63-C1.
 OS Leishmania mexicana.
 OC Eukaryota; Euzoenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 NC NCB1_Taxid=5665;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MNYC/BZ/62/M379;
 RX MEDLINE=93149206; PubMed=8426614;
 RA Medina-Acosta E., Karesa R.E., Russell D.G.;
 RT "Structurally distinct genes for the surface protease of Leishmania mexicana are developmentally regulated.";
 RL Mol. Biochem. Parasitol. 57:31-46(1993).
 CC -1- FUNCTION: Has an integral role during the infection of macrophages in the mammalian host.
 CC -1- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and P1' and basic residues at P2 and P3'. A model nonapeptide is cleaved at -Ala-Tyr-Ileu-Lys-Lys-.
 CC -1- COFACTOR: Binds 1 zinc ion per subunit (by similarity).
 CC -1- DEVELOPMENTAL STAGE: Expressed in both the promastigote and the amastigote forms.
 CC -1- SIMILARITY: Belongs to peptidase family M8.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: X64394; CAA45733.1; -.
 DR PIR: S19916; S19916.
 DR HSSP: P08148; 1LM1.
 DR MEROPS: M08.001; -.
 DR GLYCOSULEDB: P43150; -.
 DR InterPro: IPR006025; Pept_M_Zn_BS.
 DR InterPro: IPR001577; Peptidase_M8.
 DR Pfam: PF01457; Peptidase_M8; 1.
 DR PRINTS: PR00782; LSHMANOLYSIN.
 DR PROSITE: PS00142; ZINC_PROTEASE; 1.
 DR KMW Hydrolyase; Metalloprotease; Glycoprotein; Metal-binding; Zinc; Kymogen; Signal; Cell adhesion; Multigene family.
 KW SIGNAL
 FT 1 39
 FT PROPEP 40 102
 FT CHAIN 103 646
 FT METAL 266 266
 FT ACT SITE 267 267
 FT METAL 270 270
 FT METAL 336 336
 FT DISULFID 127 144
 FT DISULFID 193 232
 FT DISULFID 316 388
 FT DISULFID 395 458
 FT DISULFID 408 427
 FT DISULFID 417 492
 FT DISULFID 469 513
 FT DISULFID 518 568
 FT DISULFID 538 561
 FT CARBOHYD 86 86
 FT CARBOHYD 297 297
 FT CARBOHYD 399 399
 FT CARBOHYD 409 409
 FT CARBOHYD 433 433
 FT CARBOHYD 445 445
 FT CARBOHYD 466 466
 FT CARBOHYD 501 501
 SQ SEQUENCE 646 AA; 69054 MW; FE448DDC78C10B0A CRC64;

Query Match 13.1%; Score 99.5; DB 1; Length 646;
 Best Local Similarity 30.8%; Pred. No. 0.33;
 Matches 40; Conservative 15; Mismatches 66; Indels 9; Gaps 4;
 QY 8 AFAAIYSSSALAGVDDVITRVVTHMAHAGPDSLTSLIYQGSANALALQSDARK-- 65
 DB 241 AVGVINIPANIASRRDQVTRVTHMAHAGVSGTF---FGAVGIVQEVPLRRKDF 296
 QY 66 SETTIGSGVGNAGADVGGADN-STIELTONGFRNNA--TIDQWNAKNSDITVGYGNN 122
 DB 297 NVSVITSTVYAKARQYGCNSLEYIEIDQGAGSAGSHIKRNKDELMPAASAGY 356
 QY 123 AALVQNTASD 132
 DB 357 TALTMAVFQD 366
 RESULT 10
 OMPB_RICJA
 ID OMPB_RICJA STANDARD; PRT; 1656 AA.
 AC 006653;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Outer membrane protein B precursor (168 kDa surface-layer protein) (Surface protein antigen) (Cell surface antigen 5) (Scas) (rmpB)
 DE (rmp B) [contains: 120 kDa surface-exposed protein (Surface protein antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptidol].
 GN OMPB.
 OS Rickettsia japonica.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Rickettsiaceae; Rickettsiales; Rickettsia.
 OC NCB1_Taxid=35790;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=YH;
 RA Uchiyama T.;
 RT "Sequencing of the gene encoding the protein rmp B of Rickettsia japonica."
 RL Submitted (MAY-1997) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: THE 120 kDa SURFACE-EXPOSED PROTEIN IS A MAJOR STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (By similarity).
 CC -1- FUNCTION: THE 32 kDa BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-layer with hexagonal symmetry.
 CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAL OMPA/OMPB FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: AB003681; BAA20138.1; -.
 DR InterPro: IPR006315; Autotransport.
 DR InterPro: IPR005546; Autotransporter.
 DR Pfam: PF03797; Autotransporter_1.
 DR TIGRFAMs: TIGR01414; autotrans_bar1; 2.
 KW Antigen; S-layer; Cell wall.
 FT CHAIN 1 1338
 FT CHAIN 1339 1656
 FT DOMAIN 528 533
 FT POLY-GLY. 32 kDa BETA PEPTIDE.
 SQ SEQUENCE 1656 AA; 168097 MW; 3132A69C9D5999F CRC64;
 Query Match 13.1%; Score 99; DB 1; Length 1656;
 Best Local Similarity 24.4%; Pred. No. 1.1;
 Matches 42; Conservative 23; Mismatches 51; Indels 56; Gaps 8;

```

QY 6 VAAFAIVSGSALAGVYDQVLTVRVTHMAHAGPDSLTIYOGSANAALQSDARK 65
DB 509 VLAAGATLTDSGA-----TITGDINGGG-----GAALGSIITLANATK 547
QY 66 SETTITQSG---YGNAGDVQGGADNSTIELTONGFRNNATID-----104
DB 548 ---TLTGGANIISANGGTINFGANGTIKLTST--QNNIYVDCLAIATDQTVGVASS 602
QY 105 QWNAKNSDI--TVGOYGGNNAL-----VNOFASSSVWVROVGGNNAT 147
DB 603 LTNMQTTLTSGTIGIIGANNNTTLGQFNIGSSKTTLNGANVALNELVIIGNNGS 654

RESULT 11
ICEN_PANAN
ID_ICEN_PANAN STANDARD; PRT: 1034 AA.
AC Q47879;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ice nucleation protein inau.
GN INAU.
OS Pantoea ananas (Erwinia uredovora).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pantoea.
OX NCBI_TaxID=553;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K01N-3;
RX MEDLINE=94264407; PubMed=7764866;
RA Michigami Y., Matabe S., Abe K., Obata H., Arai S.;
RT "Cloning and sequencing of an ice nucleation active gene of Erwinia
RT uredovora.";
RL Biosci. Biotechnol. Biochem. 58:762-764(1994).
CC -1- FUNCTION: Ice nucleation proteins enable bacteria to nucleate
CC crystallization in supercooled water.
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- DOMAIN: CONTAINS IMPERFECT REPEATS OF A CONSENSUS OCTAPEPTIDE
CC A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A REGIONAL 48-RESIDUE
CC PERIODICITY IS SUPERIMPOSED.
CC -1- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE
CC NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.
CC -1- SIMILARITY: Belongs to the bacterial ice nucleation protein
CC family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D14992; BAA03636.1; -.
DR PIR; JC2143; JC2143.
DR HSSP; P06620; INA.
DR InterPro: IPR000258; Ice nucleatn.
DR Pfam: PF00818; Ice nucleation; 51.
DR PRINTS; PR00327; ICENUCLEATN.
DR PROSITE; PS00314; ICE_NUCLEATION; 34.
KW Ice nucleation; Repeat; Outer membrane.
FT DOMAIN 162 993 OCTAPEPTIDE PERIODICITY.
SQ SEQUENCE 1034 AA; 103378 MW; FA22523D333AD CRC64;

Query Match 12.9%; Score 97.5; DB 1; Length 1034;
Best Local Similarity 28.7%; Pred. No. 0.85;
Matches 43; Conservative 24; Mismatches 46; Indels 37; Gaps 11;
QY 12 IVVSGSALAGVY-DQVLTVRVTHMAHAGPDSLTIYOGSANAALQSDARKSEPTTI 70
DB 161 IATYGTSLTGHOSQLIAGIGSTETTA---GDSSTL-IAGYGTGTGA-----GSDSTL 208

```

```

QY 71 TQSGYNGADVQGGADNSTIELTONGFRNNATIDQWNAKNSDIWGOYQ-----GNNAAL 125
DB 209 V-AGYSTGTAGESSQMA-----GYGSTQT-----GMKGSDLTIG-YSGTGAQDSSSL 256
QY 126 V-----NOTASDSSVWVROVGGNNATIANQ 150
DB 257 IAGYGTGTAGESSSLT--AGYSTGTQAK 284

RESULT 12
GP63_LEIMA
ID_GP63_LEIMA STANDARD; PRT: 602 AA.
AC P08148; P15906;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 14, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Leishmanolysin precursor (EC 3.4.24.36) [Cell surface protease]
DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface
DE endopeptidase).
GN GP63.
OS Leishmania major.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 101-123.
RX MEDLINE=88154764; PubMed=3346625;
RA Button L.L., McMaster W.R.;
RT "Molecular cloning of the major surface antigen of leishmania.";
RL J. Exp. Med. 167:724-729(1988).
RN [2]
RP REVISIONS.
RA Button L.L., McMaster W.R.;
RL J. Exp. Med. 171:589-589(1990).
RN [3]
RP GPI-ANCHOR.
RX MEDLINE=91009116; PubMed=2145267;
RA Schneider P., Ferguson M.A.J., McConville M.J., Mehler A.,
RA Homans S.W., Bordier C.;
RT "Structure of the glycosyl-1-phosphatidylinositol membrane anchor of
RT the leishmania major promastigote surface protease.";
RL J. Biol. Chem. 265:16955-16964(1990).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).
RX MEDLINE=95406217; PubMed=7675788;
RA Schlagenhaut E., Etges R., Metcalf P.;
RT "Crystallization and preliminary X-ray diffraction studies of
RT leishmanolysin, the major surface metalloproteinase from Leishmania
RT major.";
RL Proteins 22:58-66(1995).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.86 ANGSTROMS).
RX MEDLINE=98416698; PubMed=9739094;
RA Schlagenhaut E., Etges R., Metcalf P.;
RT "The crystal structure of the Leishmania major surface proteinase
RT leishmanolysin.";
RL Structure 6:1035-1046(1998).
CC -1- FUNCTION: Has an integral role during the infection of macrophages
CC in the mammalian host.
CC -1- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and
CC P1' and basic residues at P2 and P3'. A model nonapeptide is
CC cleaved at Ala-Tyr-[Leu-Lys-Lys-].
CC -1- COFACTOR: Binds 1 zinc ion per subunit.
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -1- PTM: THE PHOSPHATIDYLINOSITOL MOIETY OF THE GPI-ANCHOR CONTAINS A
CC FULLY SATURATED, UNBRANCHED 1-O-ALKYL CHAIN (MAINLY C24:0) AND A
CC MIXTURE OF FULLY SATURATED UNBRANCHED 2-O-ACYL CHAINS (C12:0,
CC C14:0, C16:0, AND C18:0).
CC -1- SIMILARITY: Belongs to peptidase family M8.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: Y00647; CAA68673.1; -

DR PIR: P10221; P10221.

DR PDB: 1LMU; 17-SEP-97.

DR MEROPS: M08.001; -

DR InterPro: IPR006025; Pept M Zn BS.

DR InterPro: IPR001577; Peptidase M8.

DR Pfam: PF01457; Peptidase M8; 1.

DR PRINTS: PRO0782; LSHMANOLYSIN.

DR PROSITE: PS00142; ZINC_PROTEASE; 1.

KW Hydrolase; Metalloprotease; Glycoprotein; Metal-binding; zinc;
 KM Zymogen; Signal; Cell adhesion; GPI-anchor; 3D-structure; Lipoprotein.

FT SIGNAL 1 39
 FT PROPEP 40 100
 FT CHAIN 101 577
 FT PROPEP 578 602
 FT METAL 264 264
 FT ACT_SITE 265 265
 FT METAL 268 268
 FT METAL 334 334
 FT DISULFID 125 142
 FT DISULFID 191 230
 FT DISULFID 314 386
 FT DISULFID 393 455
 FT DISULFID 406 425
 FT DISULFID 415 489
 FT DISULFID 466 510
 FT DISULFID 515 565
 FT DISULFID 535 558
 FT CARBOHYD 300 300
 FT CARBOHYD 407 407
 FT LIPID 577 577
 FT STRAND 101 102
 FT STRAND 107 108
 FT STRAND 111 114
 FT HELIX 116 119
 FT TURN 121 122
 FT TURN 128 129
 FT STRAND 131 133
 FT STRAND 139 141
 FT HELIX 144 146
 FT HELIX 150 158
 FT TURN 159 159
 FT HELIX 160 169
 FT TURN 170 171
 FT STRAND 172 174
 FT STRAND 177 178
 FT STRAND 180 181
 FT TURN 189 190
 FT HELIX 191 193
 FT HELIX 198 202
 FT TURN 203 203
 FT STRAND 205 206
 FT STRAND 210 215
 FT TURN 221 222
 FT STRAND 226 232
 FT TURN 234 235
 FT STRAND 238 244
 FT HELIX 247 249
 FT HELIX 256 269
 FT TURN 270 271
 FT HELIX 274 279
 FT TURN 280 281
 FT STRAND 283 286
 FT HELIX 289 291
 FT STRAND 296 299
 FT HELIX 302 312
 FT TURN 313 313
 FT TURN 315 316
 FT STRAND 320 322

ACTIVATION PEPTIDE.
 LEISHMANOLYSIN.
 REMOVED IN MATURE FORM.
 ZINC (CATALYTIC).
 ZINC (CATALYTIC).
 ZINC (CATALYTIC).
 ZINC (CATALYTIC).
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 GPI-anchor amidated asparagine.

FT TURN 328 332
 FT STRAND 335 335
 FT TURN 337 339
 FT TURN 341 342
 FT STRAND 343 343
 FT TURN 344 345
 FT STRAND 353 353
 FT HELIX 356 364
 FT TURN 365 366
 FT STRAND 369 370
 FT HELIX 372 374
 FT TURN 380 383
 FT HELIX 386 390
 FT STRAND 394 395
 FT TURN 396 397
 FT STRAND 398 399
 FT TURN 402 404
 FT STRAND 413 414
 FT TURN 417 418
 FT STRAND 421 425
 FT STRAND 428 429
 FT HELIX 435 437
 FT TURN 443 444
 FT STRAND 445 446
 FT TURN 450 454
 FT STRAND 458 465
 FT TURN 466 467
 FT HELIX 470 472
 FT TURN 475 477
 FT HELIX 478 480
 FT TURN 485 486
 FT STRAND 487 494
 FT STRAND 496 496
 FT STRAND 506 516
 FT TURN 517 520
 FT STRAND 521 525
 FT TURN 527 528
 FT STRAND 533 534
 FT TURN 537 538
 FT STRAND 540 542
 FT HELIX 543 545
 FT TURN 546 546
 FT STRAND 550 550
 FT TURN 552 553
 FT STRAND 555 557
 FT HELIX 561 565
 FT TURN 566 567
 FT HELIX 569 572
 FT TURN 573 573

SEQUENCE 602 AA; 63953 MW; 982BF3245D87C43E CRC64;

Query Match 12.7%; Score 96.5; DB 1; Length 602;
 Best Local Similarity 59.5%; Pred. No. 0.56; 9; Indels 3; Gaps 1;
 Matches 22; Conservative 3; Mismatches 9;

QY 8 AFAAIVSGSALAGVYDQVTRVVTHEMAHA--SGP 41
 | : : : ||||| |||
 Db 239 AVGVINIPAMNTASRYDQVTRVVTHEMAHALGFGSP 275

RESULT 13
 ID ICEN ERWHE STANDARD; PRT; 1258 AA.
 AC P16239;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Ice nucleation protein.
 GN ICEE.
 OS Erwinia herbicola.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Pantoea.
 NX NCBI_taxid=549;

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ML;
RX MEDLINE=90152370; PubMed=2515997;
RA Warren G.U., Corotto L.V.;
RT "The consensus sequence of ice nucleation proteins from Erwinia
RL hericola, Pseudomonas fluorescens and Pseudomonas syringae.";
CC Gene 85:239-242(1989).
CC -1- FUNCTION: Ice nucleation proteins enable bacteria to nucleate
CC crystallization in supercooled water.
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- DOMAIN: CONTAINS 126 IMPERFECT REPEATS OF A CONSENSUS OCTAPEPTIDE
CC A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A REGIONAL 48-RESIDUE
CC PERIODICITY IS SUPERIMPOSED.
CC -1- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE
CC NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.
CC -1- SIMILARITY: Belongs to the bacterial ice nucleation protein
CC family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M26382; AAA24823.1; -
DR PIR; J00188; J00188.
DR HSSP; P06620; 11NA.
DR InterPro; IPR000258; Ice_nucleatn.
DR Pfam; PF00818; Ice_nucleation; 65.
DR PRINTS; PR00327; ICENUCLEATN.
DR PROSITE; PS00314; ICE_NUCLEATION; 45.
KW Ice nucleation; Repeat; Outer membrane.
FT DOMAIN 162 1217
SQ SEQUENCE 1258 AA; 125084 MW; 5908BA130077FBD4 CRC64;

Query Match 12.6%; Score 95.5; DB 1; Length 1258;
Best Local Similarity 28.7%; Pred. No. 1.6;
Matches 43; Conservative 24; Mismatches 46; Indels 37; Gaps 11;

QY 12 IVVSGSALAGYV-DQVTVRVVTHMAHSGPDSLTIYQYGSANALALQSDARKSEPTTI 70
DB 161 IATVGSSTLSTGHQSLIAGVSTETA---GDSSTL-IAVGSSTGTAGA-----DSTL 208
QY 71 TOSGVCAGADVCGAGDNSTIELTONGFPNNATTIDQWAKNSDITVQYG-----GNNAAL 125
DB 209 V-AGYGSTQTAGESSQMA-----GYGSTQT-----GKGSDDLTAAG-YGSTGTAGDSSL 256
QY 126 V-----NOTVSDSSVWVRQVFGNNAATNQ 150
DB 257 IAGYGSTQTAGESSSLT--AGYGSTQTAK 284

RESULT 14
ID ICEA_PANAN STANDARD; PRT; 1322 AA.
AC P20469;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ice nucleation protein InaA.
GN INAA.
OS Pantoea ananas (Erwinia uredovora).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pantoea.
OX NCBI_TaxID=553;
RX MEDLINE=90092494; PubMed=2599095;
RA Abe K., Watabe S., Emori Y., Watanabe M., Arai S.;
RT "An ice nucleation active gene of Erwinia ananas. Sequence similarity

```

```

RT to those of Pseudomonas species and regions required for ice
RT nucleation activity.";
RL FEBS Lett. 258:297-300(1989).
CC -1- FUNCTION: Ice nucleation proteins enable bacteria to nucleate
CC crystallization in supercooled water.
CC -1- SUBCELLULAR LOCATION: Outer membrane (By similarity).
CC -1- DOMAIN: CONTAINS MANY IMPERFECT REPEATS OF THE CONSENSUS
CC OCTAPEPTIDE A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A
CC REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.
CC -1- SIMILARITY: Belongs to the bacterial ice nucleation protein
CC family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X17316; CAA35194.1; -
DR PIR; S07053; S07053.
DR HSSP; P06620; 11NA.
DR InterPro; IPR000258; Ice_nucleatn.
DR Pfam; PF00818; Ice_nucleation; 69.
DR PRINTS; PR00327; ICENUCLEATN.
DR PROSITE; PS00314; ICE_NUCLEATION; 49.
KW Ice nucleation; Repeat; Outer membrane.
FT DOMAIN 162 1281
SQ SEQUENCE 1322 AA; 131094 MW; 89B0EE24A837039 CRC64;

Query Match 12.5%; Score 94.5; DB 1; Length 1322;
Best Local Similarity 29.3%; Pred. No. 2.1;
Matches 41; Conservative 20; Mismatches 46; Indels 33; Gaps 10;

QY 12 IVVSGSALAGYV-DQVTVRVVTHMAHSGPDSLTIYQYGSANALALQSDARKSEPTTI 70
DB 161 IATVGSSTLSTGHQSLIAGVSTETA---GDSSTL-IAVGSSTGTAGA-----GDSSTL 208
QY 71 TOSGVCAGADVCGAGDNSTIELTONGFPNNATTIDQWAKNSDITVQYG-----GNNAAL 130
DB 209 V-AGYGSTQTAGESSQMA-----GYGSTQT-----GKGSDDLTAAG-YGSTGTAGA 251
QY 131 SDSVWVRQVFGNNAATNQ 150
DB 252 ADSLSL---AGYGSTQTAGE 268

RESULT 15
ID ICEN_XANCT STANDARD; PRT; 1567 AA.
AC P18127;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ice nucleation protein.
GN INAX.
OS Xanthomonas campestris (pv. translucens).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=343;
RX MEDLINE=91080859; PubMed=2259339;
RA Zhao J., Orser C.S.;
RT "Conserved repetition in the ice nucleation gene inax from
RT Xanthomonas campestris pv. translucens.";
RL Mol. Gen. Genet. 223:163-166(1990).
CC -1- FUNCTION: Ice nucleation proteins enable bacteria to nucleate
CC crystallization in supercooled water.
CC -1- SUBCELLULAR LOCATION: Outer membrane (By similarity).
CC -1- DOMAIN: CONTAINS 153 IMPERFECT REPEATS OF THE CONSENSUS

```



```

CC      OCAPETIDE A-G-Y-G-S-T-L-T: FURTHER ON A 16-RESIDUE AND A
CC      REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.
CC      -I- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE
CC      NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.
CC      -I- SIMILARITY: Belongs to the bacterial ice nucleation protein
CC      family.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.ebi.ac.uk/announce/
CC      or send an email to license@ebi.ac.uk).
CC      -----
DR      EMBL; X52970; CAA37140.1; -.
DR      HSSP; P06620; IINA.
DR      InterPro; IPRO00258; Ice_nucleatn.
DR      Pfam; PF00818; Ice_nucleation; 81.
DR      PRINTS; PR00327; ICENUCLEATN.
DR      PROSITE; PS00314; ICE_NUCLEATION; 57.
KW      Ice nucleation; Repeat; Outer membrane.
SQ      SEQUENCE   1567 AA;  152548 MW;  C8B451D95ECAD63 CRC64;

Query Match          12.4%; Score 93.5; DB 1; Length 1567;
Query Local Similarity 29.9%; Pred. No. 3;
Matches    43; Conservative 26; Mismatches 30; Indels 45; Gaps 12;

Cc       39  SCGPSTLSIVYGSAMNALAQSDARKSETITTOSGYGN-----GADV-----GGC 84
Cc       :|||:|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Cc       1060 AGADSTL-IAGYGSTGTGA-----GSDSSLT-AGYSTQTARGSDSLTAGYSTGTGA 1109
Cc       :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Cc       85  ADNSTIE--LTQN-GFPNNATI-----DQNAKNSDITVGYG-----GNNALV----- 126
Cc       :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Cc       Db     1110 ADSLLIAGYSTQTARGYSNTIAGYGSTQTAREDSISLAG-YGSTSTAGHSDSIAGYGS 1168
Cc       :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Cc       Cc       127 NOTASDSSVMVRQYGFNGGNATAHQ 150
Cc       :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Cc       Db     1169 TOTAGYNSILT-TGYGSTQTARA 1190

```

Search completed: March 11, 2004, 18:34:55
Job time : 7.3 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Comogen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:23:54 ; Search time 30.5 Seconds
(without alignments)
1562.074 Million cell updates/sec

Title: US-09-543-407-16

Perfect score: 757

Sequence: 1 MLLKVAAPAAIVSGSALA.....DSSVMVROYFGNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPREMBL_25.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_virus.*
16: sp_bacteriap.*
17: sp_archesp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	649	85.7	152	2	033802
2	563.5	74.4	150	2	07X243
3	524	69.2	149	2	07X240
4	471.5	62.3	152	16	08CW63
5	454.5	60.0	150	2	07X237
6	340	44.9	76	2	054069
7	133	17.6	502	16	08EIH4
8	120	15.9	160	16	08CW64
9	120	15.9	160	16	083R07
10	118.5	15.7	151	16	07UC21
11	118.5	15.7	153	16	089J16
12	116	15.3	171	16	089J13
13	115	15.2	139	16	08EIH3
14	110	14.5	130	16	089J14
15	108.5	14.3	151	2	07X244
16	107.5	14.2	154	16	089J15

17	105.5	13.9	91	2	09S3J8	09S3J8 escherichia
18	102.5	13.5	152	2	07X241	07X241 citrobacter
19	101.5	13.4	2087	5	08WVW7	08WVW7 naegleria g
20	101	13.3	599	5	025273	025273 leishmania
21	99.5	13.1	582	16	P71868	P71868 mycobacteri
22	99.5	13.1	582	16	Q7TW98	Q7TW98 mycobacteri
23	99.5	13.1	641	5	08MN20	08MN20 leishmania
24	99.5	13.1	657	5	08MN21	08MN21 leishmania
25	99	13.1	1613	2	0840U5	0840U5 rickettsia
26	99	13.1	1618	2	09KXB1	09KXB1 rickettsia
27	98.5	13.0	151	2	07X238	07X238 enterobacte
28	98	12.9	145	16	Q8U6N9	Q8U6N9 agrobacteri
29	98	12.9	157	16	Q8RHG0	Q8RHG0 pseudomonas
30	97	12.8	1612	2	0840U6	0840U6 rickettsia
31	97	12.8	1617	2	07X5N9	07X5N9 rickettsia
32	96.5	12.7	348	13	093397	093397 cyprinus ca
33	96.5	12.7	644	5	043994	043994 leishmania
34	96.5	12.7	1306	2	093N36	093N36 pantoea ana
35	96.5	12.7	1341	16	Q8ED31	Q8ED31 shewanella
36	96.5	12.7	2016	5	09BIT0	09BIT0 plectreurya
37	96	12.7	598	5	025275	025275 leishmania
38	96	12.7	641	5	08MN22	08MN22 leishmania
39	96	12.7	641	5	08MN48	08MN48 leishmania
40	96	12.7	641	5	08MN23	08MN23 leishmania
41	96	12.7	641	5	08MN24	08MN24 leishmania
42	96	12.7	641	5	08MN25	08MN25 leishmania
43	95.5	12.6	179	2	033801	033801 salmonella
44	95.5	12.6	617	5	094593	094593 leishmania
45	95.5	12.6	1613	2	Q9KKB2	Q9KKB2 israeli titc

ALIGNMENTS

RESULT 1

ID	033802	PRELIMINARY;	PRT;	152 AA.
AC	033802;			
DT	01-JAN-1998 (TREMBLrel. 05, Created)			
DT	01-JAN-1998 (TREMBLrel. 05, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE	Agfa protein (Fragment).			
GN	Agfa.			
OS	Salmonella typhimurium.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OC	Enterobacteriaceae; Salmonella.			
OX	NCBI_Taxid=602;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=98053981; PubMed=9393832;			
RA	Sukupolvi S.S., Lorentz R.G., Gordon J.I., Bian Z., Pfeiffer J.D.,			
RA	Normark S.J., Rhen M.;			
RT	"Expression of thin, aggregative fimbriae promotes interaction of			
RT	Salmonella typhimurium SR-11 with mouse small intestinal epithelial			
RT	cells."			
RL	Infect. Immun. 65:5320-5325(1997).			
DR	EMBL; AJ000514; CAA04151.1; -			
FT	NON_TBR 152 152			
SO	SEQUENCE 152 AA; 15401 MW; 9DA7DADC264B006 CRC64;			

Query Match 85.7%; Score 649; DB 2; Length 152;
Best Local Similarity 88.7%; Pred. No. 1.1e-44;
Matches 134; Conservative 4; Mismatches 13; Indels 0; Gaps 0;

QY	1	MLLKVAAPAAIVSGSALAGYVDLVRYVTHEMAHSGPSTSIYQGSANALALQ	60
DB	1	MLLKVAAPAAIVSGSAVAVPQWGGGNGNGNSGPOSTSIYQGSANALALQ	60
QY	61	SDARKSETTITGSGYGADVGQAGDNSTIELTONGFRNNATIDQMANKSDITYQYGG	120
DB	61	SDARKSETTITGSGYGADVGQAGDNSTIELTONGFRNNATIDQMANKSDITYQYGG	120
QY	121	NNALVNOTASDSSVMVROYFGNNATANQY	151

DB 121 NNAALVNOTASDSSVWVROVGFNNATANOY 151

RESULT 2

Q7X243 PRELIMINARY; PRT; 150 AA.

AC Q7X243; 01-OCT-2003 (TREMBlrel. 25, Created)

DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)

DE Curlin-csgA protein.

GN CSGA.

OS Citrobacter sp. fec2.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Citrobacter.

OX NCBI_TaxID=213763;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=fec2;

RA Zogaj X., Bokranz W., Nimtz M., Romling U.;

RT "Production of Cellulose and Curli Fimbriae by Members of the Family Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";

RL Infect. Immun. 72:4151-4158 (2003).

DR EMBL; AJ515700; CAD56672.1; -.

SQ SEQUENCE 150 AA; 15016 MW; 1D71418D6973DC6 CRC64;

Query Match 74.4%; Score 563.5; DB 2; Length 150;

Best Local Similarity 74.2%; Pred. No. 8e-38;

Matches 118; Conservative 11; Mismatches 13; Indels 17; Gaps 2;

QY 1 MKLLKVAAPFAAIVVSGSALAGVYDQLVTRVVTHEM-----AHASGPDSTLSIYOGS 52

DB 1 MKLLQVAFAAIVVSGSALAGSVPO-----WGGGGGGGGSSSGSPSTLSIYOGSV 51

QY 53 ANAALATOSDARKSETTTTOSGNGADVGGAGNSTLTETONGFRNNATIDOWNAKNSD 112

DB 52 NNAALALQSDARKSDTTTHONGFNGADVGGGSDNSTIDITONGFKNNATIDONGKNSD 111

QY 113 ITVGOYGGNNALVNOTASDSSVWVROVGFNNATANOY 151

DB 112 ITVGOYGGNNALVNOTASDSSVWVROVGFNNATANOY 150

RESULT 3

Q7X240 PRELIMINARY; PRT; 149 AA.

AC Q7X240; 01-OCT-2003 (TREMBlrel. 25, Created)

DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)

DE Curlin-csgA protein.

GN CSGA.

OS Citrobacter freundii.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Citrobacter.

OX NCBI_TaxID=546;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=fec4;

RA Zogaj X., Bokranz W., Nimtz M., Romling U.;

RT "Production of Cellulose and Curli Fimbriae by Members of the Family Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";

RL Infect. Immun. 72:4151-4158 (2003).

DR EMBL; AJ515701; CAD56675.1; -.

SQ SEQUENCE 149 AA; 15260 MW; 946DD52017F648FD CRC64;

Query Match 69.2%; Score 524; DB 2; Length 149;

Best Local Similarity 72.8%; Pred. No. 1.2e-34;

Matches 110; Conservative 14; Mismatches 25; Indels 2; Gaps 1;

QY 1 MKLLKVAAPFAAIVVSGSALAGVYDQLVTRVVTHEMHAASGPDSTLSIYOGSANAALAQ 60

DB 1 MKLLKVAAPFAAIVVSGSALAGVYDQVPGNGN--HHGGSSNYGPDSSLSIYOGSNNANALQ 58

QY 61 SDARKSETTTTOSGNGADVGGAGDNSTLTETONGFRNNATIDOWNAKNSDITVGOYG 120

DB 59 SDARKSDVTTTQGRNGAVVGGADSDTSLKQTGFQNSATIDOWNAKNADISVTFQFG 118

QY 121 NNAALVNOTASDSSVWVROVGFNNATANOY 151

DB 119 RKGALVNOTASDSSVNLIQVGFNNATANOY 149

RESULT 4

Q8CW63 PRELIMINARY; PRT; 152 AA.

AC Q8CW63; 01-MAR-2003 (TREMBlrel. 23, Created)

DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)

DE Major curlin subunit precursor.

GN CSGA OR C1306.

OS Escherichia coli O6.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Escherichia.

OX NCBI_TaxID=217992;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=O6.H1 / CFT073 / ATCC 700928;

RX MEDLINE=22388234; PubMed=12471157;

RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesech P.,

RA Rasko D., Buckland E.L., Lion S.-R., Boutin A., Hackett J., Stroud D.,

RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,

RT Mobley H.L.T., Donnenberg M.S., Blattner F.R.;

RT "Extensive mosaic structure revealed by the complete genome sequence of uropathogenic Escherichia coli.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).

DR EMBL; AE016759; AAN79779.1; -.

KW Complete proteome.

SQ SEQUENCE 152 AA; 15064 MW; 3BA57F34C1240E83 CRC64;

Query Match 62.3%; Score 471.5; DB 16; Length 152;

Best Local Similarity 65.8%; Pred. No. 2e-30;

Matches 100; Conservative 18; Mismatches 33; Indels 1; Gaps 1;

QY 1 MKLLKVAAPFAAIVVSGSALAGVYDQLVTRVTH-EMAHASGPDSTLSIYOGSANAALAL 59

DB 1 MKLLKVAAPFAAIVVSGSALAGVYDQVGGGNGHGGGNGSPNELNITVGGNSALAQ 60

QY 60 QSDARKSETTTTOSGNGADVGGAGDNSTLTETONGFRNNATIDOWNAKNSDITVGOYG 119

DB 61 QADARNSDITTHOGGNGADVGGGSDSDITLORFGNSATLIDOWNKSDTWTVRQFG 120

QY 120 GNNALVNOTASDSSVWVROVGFNNATANOY 151

DB 121 GNGANAVDQVNASNSVNTVGFNNATANOY 152

RESULT 5

Q7X237 PRELIMINARY; PRT; 150 AA.

AC Q7X237; 01-OCT-2003 (TREMBlrel. 25, Created)

DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)

DE Curlin-csgA protein.

GN CSGA.

OS Enterobacter sakazakii.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Enterobacter.

OX NCBI_TaxID=28141;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=fec39;

RA Zogaj X., Bokranz W., Nimtz M., Romling U.;

RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RT Enterobacteriaceae isolated from the Human Gastrointestinal Tract."
RL Infect. Immun. 72:4151-4158(2003).
SQ EMBL; AJ515702; CAD56678.1; -.
SQ SEQUENCE 150 AA; 15112 MW; 5D8B82D872DF15F3 CRC64;
Query Match 60.0%; Score 454.5; DB 2; Length 150;
Best Local Similarity 62.1%; Pred. No. 4.6e-29;
Matches 95; Conservative 24; Mismatches 29; Indels 5; Gaps 2;
QY 1 MLLVVAFAAIVSGSALAGVYDOLVTRVTHENHA--SGPDTLSIYOGSANAALA 58
DB 1 MRFIVAAALAAIVSSSNAWGMNQ---GGMGHGHGGYGGPNSITLITYOGGNSALA 57
QY 59 IQSDARKSETTITQSGYGGADVGQADNSTIELTONGFRNNATTIDQNAKNSDITVGY 118
DB 58 IGTDAKNSVTLNISQYCGGAGADVGGQSDSSINLTNGNSATLTDQNAKSDSVNVSQY 117
QY 119 GGNNAALVNOTASDSSVMVROYGFGNNATANOY 151
DB 118 GGALNGALVDQTASNSTVNTQIGFGNHTATNOY 150
RESULT 6
ID Q54069 PRELIMINARY; PRT; 76 AA.
AC Q54069;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2003 (TREMBLrel. 19, Last annotation update)
DE SEF17 fimbria (Fragment).
GN AGPA.
OS Salmonella enteritidis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OK NCBI_TaxID=592;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SE30;
RA Cox J.M., Eglezos S., Woolcock J.B.;
RT "Virulence of Salmonella enteritidis in chickens correlates with
RT colony morphology and expression of SEF17 fimbriae."
RL Submitted (Apr-1996) to the EMBL/Genbank/DBD databases.
DR EMBL; U53207; AAA98671.1; -.
FT NON TER 1 76
FT NON TER 1 76
SQ SEQUENCE 76 AA; 7704 MW; 2FD5411241A7BCB1 CRC64;
Query Match 44.9%; Score 340; DB 2; Length 76;
Best Local Similarity 97.1%; Pred. No. 3.1e-20;
Matches 67; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 37 HASGPDSTLSIYOGSANAALALQSDARKSETTITQSGYGGADVGQADNSTIELTONG 96
DB 8 NSSGPDSTLSIYOGSANAALALQSDARKSETTITQSGYGGADVGQADNSTIELTONG 67
QY 97 FRNNATTIDQ 105
DB 68 FRNNATTIDQ 76
RESULT 7
ID Q8EIH4 PRELIMINARY; PRT; 502 AA.
AC Q8EIH4;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Conserved hypothetical protein.
GN S00865.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.

OK NCBI_TaxID=70863;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WR-1;
RX MEDLINE=22297686; PubMed=12368813;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Mehe B., Clayton R.A.,
RA Meyer T., Tsaplin A., Scott J., Beaman M., Brinkac L., Daugherty S.,
RA Deboy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
RA Vamathevan J., Weidman J., Imprial M., Lee K., Berry K., Lee C.,
RA Mueller J., Khoult H., Gill J., Uterback T.R., McDonald L.A.,
RA Feldblyum T.V., Smith H.O., Venter J.C., Nealon K.H., Fraser C.M.;
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis."
RL Nat. Biotechnol. 20:1118-1123(2002).
DR EMBL; AE015532; AAN53941.1; -.
DR TIGR; S00865;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 502 AA; 52441 MW; D08CA23D6C46B62D CRC64;
Query Match 17.6%; Score 133; DB 16; Length 502;
Best Local Similarity 29.6%; Pred. No. 0.013;
Matches 37; Conservative 21; Mismatches 59; Indels 8; Gaps 3;
QY 32 THEMASGPDSTLSIYOGSANA--ALALQSDARKSETTITQSGYGGADVGQADNS 88
DB 258 TAYLSMTTGDDNTVDVITQDGDSTNTVGDLSLADIQGDNDNTIKQKGSNGAFQVWGSDN 317
QY 89 TIELTONGFRNNATTIDQNAKNSDITVGYGGNNAALVNOTASDSSVMVROYGFGN---- 144
DB 318 DVDLRQDDANPATGAYGATDN-DPLSSKGNBELVAPATGDNSEIISQGDANFAYV 376
QY 145 NATAN 149
DB 377 DATGN 381
RESULT 8
ID Q8CW64 PRELIMINARY; PRT; 160 AA.
AC Q8CW64;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Minor curliin subunit precursor.
GN CSGB OR C1305.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OK NCBI_TaxID=217992;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O6:HI / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Raske D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli."
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
DR EMBL; AB016759; AAN79778.1; -.
KW Complete proteome.
SQ SEQUENCE 160 AA; 16963 MW; 49F68448D979B986 CRC64;
Query Match 15.9%; Score 120; DB 16; Length 160;
Best Local Similarity 28.7%; Pred. No. 0.036;
Matches 41; Conservative 17; Mismatches 67; Indels 18; Gaps 3;
QY 22 VYDQ-----LVTRVTHEMAHSGPDSTLSIYOGSANAALALQSDARKSRT 68
DB 1 MTDVQGDNNKKNKLLFMULTITIGAPGIAAAGYDLANSBYNF---AVNELSKSPNOA 56

QY 69 TTOSGNGADVGAGDNSTIELTONGFNNATIDQWNAKSDITVGYGNNALVNO 128
 DB 57 IIGAGTNNNSAQLRQGGSKLLTVAAGSSNRAKIDQTGDPYNL-AYIDAGSANDASISQ 115
 QY 129 TASDSVWVRQVGFNNATANOY 151
 DB 116 GAYGNTAMTIQKSGNNRANITQY 138

RESULT 9

Q83RUT7 PRELIMINARY; PRT; 160 AA.
 AC Q83RUT7;
 DT 01-JUN-2003 (TREMBlrel. 24, Created)
 DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
 DE Minor curlin subunit precursor, similar to CsgA.
 GN CsgB OR SF1035.
 OS Shigella flexneri.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Shigella.
 OX NCBI_TaxId=623;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=301 / Serotype 2a;
 RX MEDLINE=22272406; PubMed=12384590;
 RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
 RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
 RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
 RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
 RA Yu J.,
 RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
 through comparison with genomes of Escherichia coli K12 and O157.";
 RL Nucleic Acids Res. 30:4432-4441 (2002).
 DR EMBL; AF015131; AAN42658.1; -.
 KW Complete proteome.
 SQ SEQUENCE 160 AA; 16919 MW; 50269F5268D2A32P CRC64;

Query Match 15.9%; Score 120; DB 16; Length 160;
 Best Local Similarity 28.7%; Pred. No. 0.036;
 Matches 41; Conservative 17; Mismatches 67; Indels 18; Gaps 3;

QY 22 VYDQ-----LVTVVTHMAHSGPSTSTIYQSGANALALQSDARKSET 68
 DB 1 MYDVGQDNMKKLLFMFLITLGAAPGIAAAGYDLANSEYNF---AVNELSKSFPQAA 56
 QY 69 TTOSGNGADVGAGDNSTIELTONGFNNATIDQWNAKSDITVGYGNNALVNO 128
 DB 57 IIGAGTNNNSAQLRQGGSKLLTVAAGSSNRAKIDQTGDPYNL-AYIDAGSANDASISQ 115
 QY 129 TASDSVWVRQVGFNNATANOY 151
 DB 116 GAYGNTAMTIQKSGNNRANITQY 138

RESULT 10

Q7UCZ1 PRELIMINARY; PRT; 151 AA.
 AC Q7UCZ1;
 DT 01-OCT-2003 (TREMBlrel. 25, Created)
 DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
 DE Minor curlin subunit.
 GN CsgB OR S1108.
 OS Shigella flexneri.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Shigella.
 OX NCBI_TaxId=623;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=2457T / ATCC 700930 / Serotype 2a;
 RX MEDLINE=22590274; PubMed=12704152;

RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
 RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
 RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
 RA Schwartz D.C., Blattner F.R.,
 RT "Complete genome sequence and comparative genomics of Shigella
 RT flexneri serotype 2a strain 2457T".
 RL Infect. Immun. 71:2775-2786 (2003).
 DR EMBL; AF016981; AAP16542.1; -.
 KW Complete proteome.
 SQ SEQUENCE 151 AA; 15868 MW; 5D5D266B964014A0 CRC64;

Query Match 15.7%; Score 118.5; DB 16; Length 151;
 Best Local Similarity 31.6%; Pred. No. 0.045;
 Matches 37; Conservative 13; Mismatches 62; Indels 5; Gaps 2;

QY 35 MAHSGPDSLTSTIYQSGANALALQSDARKSETTITOSGNGADVGAGDNSTIELTQ 94
 DB 18 IAAAGYDLANSEYNF---AVNELSKSFPQAAIIGAGTNNNSAQLRQGGSKLLAVAAQ 73
 QY 95 NGFRNATIDQWNAKSDITVGYGNNALVNO TASDSVWVRQVGFNNATANOY 151
 DB 74 EGSSNRAKIDQTGDPYNL-AYIDAGSANDASISQ GAYGNTAMTIQKSGNNRANITQY 129

RESULT 11

Q89J16 PRELIMINARY; PRT; 153 AA.
 AC Q89J16;
 DT 01-JUN-2003 (TREMBlrel. 24, Created)
 DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
 DE CsgB protein.
 GN CsgB OR BL15297.
 OS Bradyrhizobium japonicum.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Bradyrhizobium; Bradyrhizobium.
 OX NCBI_TaxId=375;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=USDA 110;
 RX MEDLINE=2248498; PubMed=12597275;
 RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiyama T.,
 RA Sasamoto S., Watanabe A., Idegawa K., Iriguchi M., Kawashima K.,
 RA Kohara M., Matsunoto M., Shimpo S., Tsunokawa H., Wada T., Yamada M.,
 RA Tabata S.,
 RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
 RT Bradyrhizobium japonicum USDA110.";
 RL DNA Res. 9:189-197 (2002).
 DR EMBL; AP005954; BAC50562.1; -.
 KW Complete proteome.
 SQ SEQUENCE 153 AA; 15991 MW; 4CE71DEAC375145B CRC64;

Query Match 15.7%; Score 118.5; DB 16; Length 153;
 Best Local Similarity 27.0%; Pred. No. 0.046;
 Matches 41; Conservative 32; Mismatches 56; Indels 23; Gaps 6;

QY 2 KLIKVAA-FAIVVSGSALAGVYDQLVTVVTHMAHSGPSTSTIYQSGANALALQ 59
 DB 10 RYLAVALLAIAIGATQASAGSIQSGVT-----NPNVISITITQFGANDVQPVIT 58
 QY 60 QSDARKSETTITOSGNGADVGAGDNSTIELTONGFNNATIDQWNAKSDITVGYG 119
 DB 59 EENSRRVIAVIG-----GSGTVDAIT--IQGTBNYANVIOGSG-TTNAAVGQSG 108
 QY 120 GNNAAVNO TASDSVWVRQVGFNNATANOY 151
 DB 109 LSWTADITQIGNSTNALLIQIGMNSGAVRQF 140

RESULT 12

Q89J13 PRELIMINARY; PRT; 171 AA.
 AC Q89J13;
 DT 01-JUN-2003 (TREMBlrel. 24, Created)

DT 01-JUN-2003 (T-EMBLrel. 24, last sequence update)
 DT 01-JUN-2003 (T-EMBLrel. 24, last annotation update)
 DE CsgA protein.
 CS CsgA OR BL5300.
 OS Bradyrhizobium japonicum.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Bradyrhizobiaceae; Bradyrhizobium.
 NC NCB1_Taxid=375;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=USDA 110;
 RX MEDLINE=2248498; PubMed=12597275;
 RA Kaneo T., Nakamura Y., Sato S., Minamisawa K., Uchiyumi T.,
 RA Sasamoto S., Watanabe A., Ideasa K., Iriguchi M., Kawashima K.,
 RA Kohara M., Matsumoto M., Shimpō S., Tsuruoka H., Wada T., Yamada M.,
 RA Tabata S.;
 RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
 RT Bradyrhizobium japonicum USDA110.";
 RL DNA Res. 9:189-197(2002).
 DR EMBL; AP005954; BAC50565.1; -.
 KW Complete proteome.
 SQ SEQUENCE 171 AA; 17448 MW; 995D08C01498381 CRC64;

Query Match 15.3%; Score 116; DB 16; Length 171;
 Best Local Similarity 37.5%; Pred. No. 0.083;
 Matches 36; Conservative 10; Mismatches 40; Indels 10; Gaps 3;

QY 55 AALALQSDARKSETTT-OSGYGNGADVQ-GADNSTIELTONGFRNNATIDQWAKND 112
 DB 49 AVALASAAQAMANTSTTVQVGLVNGSSVTQNGCLTNDSSSTTGILNGASTMQGTSSPS- 107
 QY 113 ITVGYGNNALVNOTASDSSVMVQVFGNNATA 148
 DB 108 -----LNNVSTVNQAGVNSATTTQVAFGNNGSA 136

RESULT 13
 ID 08E1H3 PRELIMINARY; PRT; 139 AA.

AC 08E1H3;
 DT 01-MAR-2003 (T-EMBLrel. 23, Created)
 DT 01-MAR-2003 (T-EMBLrel. 23, last sequence update)
 DT 01-MAR-2003 (T-EMBLrel. 23, last annotation update)
 DE Minor curlin subunit CsgB, putative.
 GN SC00866.
 OS Shewanella oneidensis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
 OC Alteromonadaceae; Shewanella.
 NC NCB1_Taxid=70863;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=MR-1;
 RX MEDLINE=22297686; PubMed=12368813;
 RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
 RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
 RA Meyer T., Tsapin A., Scott J., Beanan M., Binkac L., Daugherty S.,
 RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
 RA Madupu R., Peterson R.J., Umayam L.A., White O., Wolf A.M.,
 RA Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,
 RA Mueller J., Khouri H., Gill J., Usterback T.R., McDonald L.A.,
 RA Feldlyum T.V., Smith H.O., Venter J.C., Neilson K.H., Fraser C.M.;
 RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
 RT Shewanella oneidensis.";
 RL Nat. Biotechnol. 20:118-123(2002).
 DR EMBL; AEO15532; AAN53942.1; -.
 DR TIGR; SC00866; -.
 KW Complete proteome.
 SQ SEQUENCE 139 AA; 14811 MW; 41EC1CFA76957920 CRC64;

Query Match 15.2%; Score 115; DB 16; Length 139;
 Best Local Similarity 27.0%; Pred. No. 0.077;
 Matches 33; Conservative 23; Mismatches 52; Indels 14; Gaps 3;

QY 30 VVTHEMAHSGPDSITLIYQGSANAALQSDARKSETTTQSGYGNGADVQGADNST 89
 DB 32 ITLQALIERSGRDNIDLVQOGTANQIVFQSGDNS-AVYVQAGDNLISLTQIGTNE 90
 QY 90 IELTONGFRNNATIDQWAKNSDITVGYGNGNNALVNOTASDSSVMVQVFGNNATAN 149
 DB 91 VOLLOVGAQNKASIT-----QIGNDNLVQNLQGS-GNFSIQIADGALISIT 137
 QY 150 QY 151
 DB 138 QY 139

RESULT 14
 ID 089J14 PRELIMINARY; PRT; 130 AA.

AC 089J14;
 DT 01-JUN-2003 (T-EMBLrel. 24, Created)
 DT 01-JUN-2003 (T-EMBLrel. 24, last sequence update)
 DE B115299 protein.
 GN B115299.
 OS Bradyrhizobium japonicum.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Bradyrhizobiaceae; Bradyrhizobium.
 NC NCB1_Taxid=375;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=USDA 110;
 RX MEDLINE=2248498; PubMed=12597275;
 RA Kaneo T., Nakamura Y., Sato S., Minamisawa K., Uchiyumi T.,
 RA Sasamoto S., Watanabe A., Ideasa K., Iriguchi M., Kawashima K.,
 RA Kohara M., Matsumoto M., Shimpō S., Tsuruoka H., Wada T., Yamada M.,
 RA Tabata S.;
 RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
 RT Bradyrhizobium japonicum USDA110.";
 RL DNA Res. 9:189-197(2002).
 DR EMBL; AP005954; BAC50564.1; -.
 KW Complete proteome.
 SQ SEQUENCE 130 AA; 12699 MW; ACFB2D66A48D260F CRC64;

Query Match 14.5%; Score 110; DB 16; Length 130;
 Best Local Similarity 24.5%; Pred. No. 0.18;
 Matches 36; Conservative 29; Mismatches 62; Indels 20; Gaps 4;

QY 4 LKVAAPFAIVGSGALAGYDQLVTRVTHEMAHSGPDSITLIYQGSANAALQSDA 63
 DB 1 MKITLVATATLALSALTIVDAQ-----AGNSASVLDQFGTTNSFSISQTS 45
 QY 64 RKSETTTQSGYGNGADVQGADNSTIELTONGFRNNATIDQWAKNSDITVGYGNGNA 123
 DB 46 TSNNAFTTQFGATNTATTTQTSLLTVNTAVVG-QGGTTA---TASNTALT-GQVGGSNS 100
 QY 124 ALVNOTASDSSVMVQVFGNNATANQ 150
 DB 101 SLIGQIGANNNTAGVQQLGILNGSTILQ 127

RESULT 15
 ID 07X244 PRELIMINARY; PRT; 151 AA.

AC 07X244;
 DT 01-OCT-2003 (T-EMBLrel. 25, Created)
 DT 01-OCT-2003 (T-EMBLrel. 25, last sequence update)
 DT 01-OCT-2003 (T-EMBLrel. 25, last annotation update)
 DE Nucleation component of curlin monomers.
 GN CsgB.
 OS Citrobacter sp. fec2.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Citrobacter.
 NC NCB1_Taxid=213763;
 RN (1)
 RP SEQUENCE FROM N.A.

RC STRAIN=Fec2;
 RA Zoga J X.; Bokranz W.; Nimtz M.; Romling U.;
 RT "Production of Cellulose and Curli Fimbriae by Members of the Family
 RT Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
 RL Infect Immun. 72:4151-4158(2003).
 DR EMBL; AJ515700; CAD56671.1; -;
 SQ SEQUENCE 151 AA; 16158 MW; BD00AF57E1400704 CRC64;

Query Match 14.3%; Score 108.5; DB 2; Length 151;
 Best Local Similarity 27.4%; Pred. No. 0.29;
 Matches 32; Conservative 20; Mismatches 48; Indels 17; Gaps 3;

OY 51 GSANALALQDPARKE-----TTTQSGYNGADVGGADNSTELTQNGFR----- 98
 Db 14 GAPGIASATSYDIASHSEYVFAVNFELSKSFNQALITGVGTNNSAKMREQSKLSVVSQ 73
 OY 99 ----NNATIDQWNAKNSDITVGQYGVGNNAALVNOTASDSSVAVRQVFGNNATANQY 151
 Db 74 EGGSRAKVDOSGATNF-AYIAQSGHSDASISQSNYGNATMTIIOKSGGNKRNATITQY 129

Search completed: March 11, 2004, 18:40:14
 Job time : 32.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:13:53 ; Search time 45.9 Seconds
(without alignments)
929.514 Million cell updates/sec

Title: US-09-543-407-18

Perfect score: 780
Sequence: 1 MLLRLVAAFAAIVSGSALA.....DSSVMVROYFGNNATANY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_293a04:*
1: geneseqp19808:*
2: geneseqp19908:*
3: geneseqp20008:*
4: geneseqp20018:*
5: geneseqp20028:*
6: geneseqp20038:*
7: geneseqp20038:*
8: geneseqp20048:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	780	100.0	151	3	AAB36349
2	691	88.6	151	2	AAB74625
3	691	88.6	151	2	AAB36341
4	686	87.9	151	2	AAB23570
5	675	86.5	151	3	AAB36353
6	622	79.7	151	3	AAB36350
7	618	79.2	151	3	AAB36348
8	613	78.6	151	3	AAB36346
9	611	78.3	151	3	AAB36347
10	608	77.9	151	3	AAB36352
11	601	77.1	151	3	AAB36354
12	600	76.9	151	3	AAB36351
13	599	76.8	151	3	AAB36355
14	523	67.1	151	7	AAB36343
15	518	66.4	151	7	ABR82641
16	506	64.9	120	2	AAB62761
17	506	64.9	120	2	AAB23569
18	445	57.1	142	2	AAB52664
19	373	47.8	122	2	AAB52663
20	337	30.4	45	3	AAB36316
21	132	16.9	22	3	AAB36318
22	123	15.8	23	3	AAB36321
23	123	15.8	23	3	AAB36326
24	123	15.8	23	3	AAB36328
25	115	14.7	22	3	AAB36325

26	115	14.7	22	3	AAB36339	AAB36339
27	115	14.7	22	3	AAB36320	AAB36320
28	111	14.2	22	3	AAB36322	AAB36322
29	111	14.2	22	3	AAB36327	AAB36327
30	111	14.2	22	3	AAB36337	AAB36337
31	109	14.0	24	7	ABR82644	ABR82644
32	107.5	13.8	151	3	AAB36344	AAB36344
33	106.5	13.7	151	3	AAB36342	AAB36342
34	102	13.1	26	7	ABR82649	ABR82649
35	101	12.9	26	7	ABR82642	ABR82642
36	98	12.6	26	7	ABR82645	ABR82645
37	96	12.3	19	3	AAB36323	AAB36323
38	96	12.3	19	3	AAB36336	AAB36336
39	96	12.3	19	3	AAB36328	AAB36328
40	95	12.2	24	7	ABR82647	ABR82647
41	93.5	12.0	287	4	ABR66343	ABR66343
42	93	11.9	974	6	ABU17075	ABU17075
43	92	11.8	23	3	AAB36331	AAB36331
44	92	11.8	975	6	ADA33477	ADA33477
45	91.5	11.7	580	5	AA666008	AA666008

ALIGNMENTS

RESULT 1
AAB36349
ID AAB36349 standard; protein; 151 AA.
XX
AC AAB36349;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#4 amino acid sequence SEQ ID NO:18.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbria; epitope;
KW vaccine; immune response; immunogen.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000MO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UUVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay W;
XX
DR WPI; 2000-672631/65.
XX
N-PSDB; AAC64625.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence which encodes foreign epitope or antigen, expresses recombinant AgfA protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 136, 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a segment of the gene has been replaced by a segment of a foreign DNA sequence which encodes a foreign epitope or antigen. Also described are: (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended assembly system of strains of Salmonella, Escherichia coli and enterobacteriaceae for the production of fimbriae comprising recombinant AgfA, CsgA and AgfA-homologue fimbria subunits, respectively; (2) directing recombination of a recombinant gene into the chromosome of the homologous species; (3) directing recombination of a recombinant gene back into the chromosome of the homologous species, replacing the native

CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell) the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

Sequence 151 AA;

Query Match 100.0%; Score 780; DB 3; Length 151;
 Best Local Similarity 100.0%; Pred. No. 1.7e-67;
 Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLTKVAFPAIVVSGSALAGVVPWGGGNNNGGNSGPDYDQVTRVVTHEMAHALQ 60
 DB 1 MKLTKVAFPAIVVSGSALAGVVPWGGGNNNGGNSGPDYDQVTRVVTHEMAHALQ 60
 QY 61 SDARKSETTITGSGYNGADVGQADNSTIELTONGFRNNATIDQWNAKSDITVGYGG 60
 DB 61 SDARKSETTITGSGYNGADVGQADNSTIELTONGFRNNATIDQWNAKSDITVGYGG 120
 QY 121 NNALVNOTASDSVWVRQVGFNNATANQY 151
 DB 121 NNALVNOTASDSVWVRQVGFNNATANQY 151

RESULT 2

ID AAR74625 standard; protein; 151 AA.
 AC AAR74625;

DT 25-MAR-2003 (revised)
 DT 26-JUN-1995 (first entry)

DE Agfa sequence.

KW Salmonella; Agfa; vaccine.

OS Salmonella.

PN WO9425598-A2.

PD 10-NOV-1994.

PF 26-APR-1994; 94WO-IB0000207.

PR 26-APR-1993; 93US-00054452.

PA (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
 (KING/) KING J.

PI Kay MW, Collinson SK, Clouthier SC, Doran JL;
 DR WPI; 1994-358275/44.
 DR N-PSDB; AAG87467.

PT Eliciting an immune response to Salmonella - using attenuated Salmonella

PT strains, vector constructs, or compans. contg. fimbrial type proteins.

PS Disclosure; Fig 7B; 95pp; English.

XX The Salmonella Agfa protein and DNA are used in vaccine and genetic

CC immunization compositions, respectively, to elicit an immune response to
 CC Salmonella in animals (e.g. food producing animals) and humans. (Updated
 CC on 25-MAR-2003 to correct FN field.)
 XX Sequence 151 AA;

Query Match 88.6%; Score 691; DB 2; Length 151;
 Best Local Similarity 90.7%; Pred. No. 6.8e-59;
 Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 1 MKLTKVAFPAIVVSGSALAGVVPWGGGNNNGGNSGPDYDQVTRVVTHEMAHALQ 60
 DB 1 MKLTKVAFPAIVVSGSALAGVVPWGGGNNNGGNSGPDYDQVTRVVTHEMAHALQ 60
 QY 61 SDARKSETTITGSGYNGADVGQADNSTIELTONGFRNNATIDQWNAKSDITVGYGG 120
 DB 61 SDARKSETTITGSGYNGADVGQADNSTIELTONGFRNNATIDQWNAKSDITVGYGG 120
 QY 121 NNALVNOTASDSVWVRQVGFNNATANQY 151
 DB 121 NNALVNOTASDSVWVRQVGFNNATANQY 151

RESULT 3

ID AAB36341 standard; protein; 151 AA.
 AC AAB36341;

DT 26-FEB-2001 (first entry)

DE Salmonella enteritidis Agfa amino acid sequence SEQ ID NO:5.

KW Salmonella; agfa; Chromosomal gene replacement; fimbrial; epitope;
 vaccine; immune response; immunogen.

OS Salmonella enteritidis.

PN WO200060102-A2.

PD 12-OCT-2000.

PF 05-APR-2000; 2000WO-CA000356.

PR 05-APR-1999; 99US-0127888P.

PA (UYVI-) UNIV VICTORIA.

PI White AP, Doran JL, Collinson SK, Kay MW;
 DR WPI; 2000-672631/65.
 DR N-PSDB; AAC64617.

PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence

PT which encodes foreign epitope or antigen, expresses recombinant Agfa

PS protein useful for eliciting immune response in animal.

Disclosure; Page 135; 139pp; English.

CC The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SAF17/RAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa. CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or

CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

XX Sequence 151 AA;

Query Match 88.6%; Score 691; DB 3; Length 151;
 Best Local Similarity 90.7%; Pred. No. 6.8e-59;
 Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 1 MLLKVAAPAAIVSGSALAGVVPWGCGGNHNGGNSGPDYDQVTRVVTHEMAHALQ 60
 DB 1 MLLKVAAPAAIVSGSALAGVVPWGCGGNHNGGNSGPDSTLSIYOGSANAALALQ 60
 QY 61 SDARKSETTITGSGYNGADVQAGADNSTIELTONGFRNNATIDQNNAKNSDITVQYCG 120
 DB 61 SDARKSETTITGSGYNGADVQAGADNSTIELTONGFRNNATIDQNNAKNSDITVQYCG 120
 QY 121 NNPAALVNOTASDSSVWVROVGFNNATANQY 151
 DB 121 NNPAALVNOTASDSSVWVROVGFNNATANQY 151

RESULT 4

AAW23570 ID AAW23570 standard; protein; 151 AA.

XX AAW23570;
 XX 25-MAR-2003 (revised)
 XX 29-SEP-1997 (first entry)
 XX Salmonella enteritidis 27655-3b agfa.
 XX Enteropathogenic bacteria; enterobacteria; S. enteritidis; antibody.
 XX Salmonella enteritidis.
 XX Key Location/Qualifiers
 XX FT Misc-difference 123
 XX FT Misc-difference 123 /note= "Encoded by GCC"
 XX US5635617-A.
 XX 03-JUN-1997.
 XX 26-APR-1994; 94US-00233788.
 XX 26-APR-1993; 93US-00054452.
 XX (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
 XX Collinson SK, Kay WW, Doran JL;
 XX MPI; 1997-309886/28.
 XX N-PSDB; AAT74142.
 XX Isolated Salmonella gene agfa - used for diagnosis of Salmonella or
 XX enteropathogenic bacteria of the Enterobacteria family.
 XX Example 2; Fig 7; 85pp; English.
 XX The present sequence represents agfa encoded by the full agfa gene
 XX derived from Salmonella enteritidis 27655-3b. The nucleic acid can be

CC used to provide diagnostic assays for Salmonella and/or enteropathogenic
 CC bacteria of the family Enterobacteria. It can also be used to provide
 CC proteins and antibodies which can be used for assays. The nucleic acid
 CC sequence can be used to provide probes or primers which can specifically
 CC hybridize to nucleic acid molecules from greater than 99% of Salmonella
 CC strains that are pathogenic to warm-blooded animals relative to nucleic
 CC acid molecules from virtually all other microbial organisms. (Updated on
 CC 25-MAR-2003 to correct PR field.)

XX Sequence 151 AA;

Query Match 87.9%; Score 686; DB 2; Length 151;
 Best Local Similarity 90.1%; Pred. No. 2.1e-58;
 Matches 136; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

QY 1 MLLKVAAPAAIVSGSALAGVVPWGCGGNHNGGNSGPDYDQVTRVVTHEMAHALQ 60
 DB 1 MLLKVAAPAAIVSGSALAGVVPWGCGGNHNGGNSGPDSTLSIYOGSANAALALQ 60
 QY 61 SDARKSETTITGSGYNGADVQAGADNSTIELTONGFRNNATIDQNNAKNSDITVQYCG 120
 DB 61 SDARKSETTITGSGYNGADVQAGADNSTIELTONGFRNNATIDQNNAKNSDITVQYCG 120
 QY 121 NNPAALVNOTASDSSVWVROVGFNNATANQY 151
 DB 121 NNPAALVNOTASDSSVWVROVGFNNATANQY 151

RESULT 5

AAB36353 ID AAB36353 standard; protein; 151 AA.

XX AAB36353;
 XX 26-FEB-2001 (first entry)
 XX Agfa::PT3#8 amino acid sequence SEQ ID NO:26.
 XX Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
 XX vaccine; immune response; immunogen.
 XX Escherichia coli.
 XX Salmonella enteritidis.
 XX Synthetic.
 XX WO200060102-A2.
 XX 12-OCT-2000.
 XX 05-APR-2000; 2000WO-CA000356.
 XX 05-APR-1999; 99US-0127888P.
 XX (UYVI-) UNIV VICTORIA.
 XX White AP, Doran JL, Collinson SK, Kay WW;
 XX MPI; 2000-672631/65.
 XX N-PSDB; AAC64629.
 XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 XX PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 XX protein useful for eliciting immune response in animal.
 XX Disclosure; Page 138; 139pp; English.
 XX The present invention describes a recombinant agfa gene (1) where a
 XX segment of the gene has been replaced by a segment of a foreign DNA
 XX sequence which encodes a foreign epitope or antigen. Also described are:
 XX (1) use of thin aggregative fimbriae (SEF17/7AF) nucleation depended
 XX assembly system of strains of Salmonella, Escherichia coli and
 XX Enterobacteriaceae for the production of fimbriae comprising recombinant
 XX Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)

CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

Sequence 151 AA;

Query Match 86.5%; Score 675; DB 3; Length 151;
 Best Local Similarity 81.9%; Pred. No. 2.4e-57;
 Matches 136; Conservative 0; Mismatches 0; Indels 30; Gaps 2;

QY 1 MKLKVAFAIVSGSALAGVVPQWGGGNNHNGGNSGPD-----YDQ 45
 DB 1 MKLKVAFAIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYGSANALAQ 60
 QY 46 LVTRVVTHEMAHALQSDARKSETTITGSGNGADVGGADNSTIELTONGFRNNATTDQ 105
 DB 61 LVTRVVTHEMAHA-----GYNGADVGGADNSTIELTONGFRNNATTDQ 105
 QY 106 MNKNSDITVGYGNNALVNOTASDSSVWVRQVFGNNATANOY 151
 DB 106 MNKNSDITVGYGNNALVNOTASDSSVWVRQVFGNNATANOY 151

RESULT 6
 AAB36350
 ID AAB36350 standard; protein, 151 AA.
 XX
 AC AAB36350;
 XX
 DT 26-FEB-2001 (first entry)
 XX

DE Agfa::PT#3 amino acid sequence SEQ ID NO:20.
 XX
 KW *Salmonella*; agfa; chromosomal gene replacement; fimbrial; epitope;
 KW vaccine; immune response; immunogen.
 XX

OS *Salmonella enteritidis*.
 OS *Escherichia coli*.
 OS Synthetic.
 XX

PN WO200060102-A2.
 XX

PD 12-OCT-2000.
 XX

PF 05-APR-2000; 2000WO-CA000356.
 XX

FR 05-APR-1999; 99US-0127888P.
 XX

PA (UYVI-) UNIV VICTORIA.
 XX

PI White AP, Doran JL, Collison SK, Kay WJ;
 XX

DR WPI, 2000-672631/65.
 XX

DR N-PSDB; AAC64626.
 XX

PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 which encodes foreign epitope or antigen, expresses recombinant Agfa

PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 137; 139pp; English.
 XX

CC The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/7AP) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

Sequence 151 AA;

Query Match 79.7%; Score 622; DB 3; Length 151;
 Best Local Similarity 73.6%; Pred. No. 3.3e-52;
 Matches 128; Conservative 0; Mismatches 0; Indels 46; Gaps 2;

QY 1 MKLKVAFAIVSGSALAGVVPQWGGGNNHNGGNSGPD-----YDQ 42
 DB 1 MKLKVAFAIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYGSANALAQ 60
 QY 43 -----YDQLVTRVVTHEMAHALQSDARKSETTITGSGNGADVGGADNSTIELTONGF 97
 DB 61 SDARKYDQLVTRVVTHEMAHA-----GQGANSTIELTONGF 97
 QY 98 RNNATIDQMNKNSDITVGYGNNALVNOTASDSSVWVRQVFGNNATANOY 151
 DB 98 RNNATIDQMNKNSDITVGYGNNALVNOTASDSSVWVRQVFGNNATANOY 151

RESULT 7
 AAB36348
 ID AAB36348 standard; protein, 151 AA.
 XX
 AC AAB36348;
 XX

DT 26-FEB-2001 (first entry)
 XX

DE Agfa::PT#3 amino acid sequence SEQ ID NO:16.
 XX

KW *Salmonella*; agfa; chromosomal gene replacement; fimbrial; epitope;
 KW vaccine; immune response; immunogen.
 XX

OS *Salmonella enteritidis*.
 OS *Escherichia coli*.
 OS Synthetic.
 XX

PN WO200060102-A2.
 XX

PD 12-OCT-2000.
 XX

PF 05-APR-2000; 2000WO-CA000356.
 XX

Yy	1	MKLKVAFAAIVVSGSALAGVVPDVGCGGHNHGGNSGSDPYDOLVTRVVTTHMAHALIQ	60
Db	1	MKLKVAFAAIVVSGSALAGVVPDVGCGGHNHGGNSGSDPYDOLVTRVVTTHMAHALIQ	60

1 MKLKVAFAAIVVSGSALAGVVPDVGCGGHNHGGNSGSDPYDOLVTRVVTTHMAHALIQ 60
 1 MKLKVAFAAIVVSGSALAGVVPDVGCGGHNHGGNSGSDPYDOLVTRVVTTHMAHALIQ 60

QY 61 SPARKSETTITQSGVGNAGNDVQAGADNSTLETQNGFRNNATIDQNNAKNSDTTGVQYGG 120

Db 61 SPARKSETTITQSGVGNAGNDVQAGADNSTLETQNGFRNNATIDQNNAKNSDTTGVQYD 120

QY 121 NNAALVNCQASDSSVMVRQVGFQGNATAQY 151

Db 121 LVTRVYTHEMAHASVMVRQVGFQGNATAQY 151

RESULT 10

AAB36352

ID AAB36352 standard; protein; 151 AA.

AC AAB36352;

XX

XX 26-FEB-2001 (first entry)

DT

XX

DE AgfA:PT#7 amino acid sequence SEQ ID NO:24.

XX

XX Salmonella; agfA; chromosomal gene replacement; fimbria; epitope;

KW vaccine; immune response; immunogen.

XX

XX Salmonella enteritidis.

OS Escherichia coli.

OS Synthetic.

OS

PN WO200060102-A2.

PN

PD 12-OCT-2000.

PD

PF 05-APR-2000; 2000WO-CA000356.

PF

PR 05-APR-1999; 99US-0127888P.

PR

PA (UYVI-) UNIV VICTORIA.

XX

PI White AP, Doran JL, Collison SK, Kay WW;

XX

DR WPI; 2000-672631/65.

DR N-PSDB; AAC64628.

PT

PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence

PT which encodes foreign epitope or antigen, expresses recombinant AgfA

PT protein useful for eliciting immune response in animal.

XX

PS Disclosure; Page 138; 139pp; English.

XX

XX The present invention describes a recombinant agfA gene (I) where a

CC segment of the gene has been replaced by a segment of a foreign DNA

CC sequence which encodes a foreign epitope or antigen. Also described are:

CC (1) use of thin aggregative fimbriae (SBE17/TFAP) nucleation depended

CC assembly system of strains of Salmonella, Escherichia coli and

CC Enterobacteriaceae for the production of fimbriae comprising recombinant

CC AgfA, CSga and AgfA-homologue fimbria subunits, respectively; (2)

CC directing recombination of a recombinant gene into the chromosome of the

CC homologous species; (3) directing recombination of a recombinant gene

CC back into the chromosome of the homologous species, replacing the native

CC copy of that gene; and (4) eliciting an immune response in an animal,

CC comprising separating an amino acid polymer comprising a recombinant AgfA

CC protein containing a replacement segment or segments of foreign amino

CC acid sequence or sequences grown on a Salmonella, E. coli or

CC Enterobacteriaceae host cell, from the host cell and introducing the

CC polymer into the animal in conjunction with a carrier or diluent. (I) is

CC useful for the expression of recombinant AgfA protein which is useful for

CC eliciting an immune response in an animal. In a fimbrial presentation

CC system the heterologous antigens are presented in high numbers (up to

CC 500,000 copies/cell), the hybrid fimbria protein possesses both the

CC immunogenicity and adhesion properties relevant for an efficient live

CC vaccine, the carrier fimbrial subunit proteins are usually strong

CC immunogens, which may be important for directing an immune response

CC against the inserted epitope, and hybrid fimbriae are easy and

CC inexpensive to purify in large amount. The present sequence is given in

CC (1) use of thin aggregative fimbriae (SEF17/7AF) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

SO Sequence 151 AA;

Query Match 76.9%; Score 600; DB 3; Length 151;
 Best Local Similarity 81.5%; Pred. No. 4.4e-50;
 Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;

QY 1 MLLTVAAFAAIVVSGSALAGVVPQWGGGHNHNGSGSPDYDQVTVRVVTHEMAHALQ 60
 Db 1 MLLTVAAFAAIVVSGSALAGVVPQWGGGHNHNGSGSPDSTLSIYQGSANAALALQ 60
 QY 61 SDARSETTITQSGYGNAGADVQAGADNSTIELTONGFRNNAATTDQWNAKNSDITVQGYG 120
 Db 61 SDARSETTITQSGYGNAGADVQAGADNSTIELTONGFRNNAATTDQVTVRVTHEMAHALQ 120
 QY 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151
 Db 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151

RESULT 13
 AAB36355
 ID AAB36355 standard; protein, 151 AA.

XX AAB36355;

DT 26-FEB-2001 (first entry)

DE Agfa::PT3#10 amino acid sequence SEQ ID NO:30.

KM *Salmonella*: agfa; chromosomal gene replacement; fimbrial; epitope;

OS *Escherichia coli*.

OS Synthetic.

PN WO200060102-A2.

PD 12-OCT-2000.

PF 05-APR-2000; 2000MO-CA000356.

PR 05-APR-1999; 99US-0127888P.

PA (UYVI-) UNIV VICTORIA.

PI White AP, Doran JL, Collison SK, Kay WM;
 XX WPI, 2000-672631/65.

DR N-PSDB; AAC64631.
 XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 139; 139p; English.

CC The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/7AF) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

SO Sequence 151 AA;

Query Match 76.8%; Score 599; DB 3; Length 151;
 Best Local Similarity 81.5%; Pred. No. 5.5e-50;
 Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;

QY 1 MLLTVAAFAAIVVSGSALAGVVPQWGGGHNHNGSGSPDYDQVTVRVVTHEMAHALQ 60
 Db 1 MLLTVAAFAAIVVSGSALAGVVPQWGGGHNHNGSGSPDSTLSIYQGSANAALALQ 60
 QY 61 SDARSETTITQSGYGNAGADVQAGADNSTIELTONGFRNNAATTDQWNAKNSDITVQGYG 120
 Db 61 SDARSETTITQSGYGNAGADVQAGADNSTIELTONGFRNNAATTDQVTVRVTHEMAHALQ 120
 QY 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151
 Db 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151

RESULT 14
 AAB36343
 ID AAB36343 standard; protein, 151 AA.

XX AAB36343;

DT 26-FEB-2001 (first entry)

DE *Escherichia coli* CsgA amino acid sequence SEQ ID NO:7.

KM *Salmonella*: agfa; chromosomal gene replacement; fimbrial; epitope;

OS *Escherichia coli*.

PN WO200060102-A2.

PD 12-OCT-2000.

PF 05-APR-2000; 2000MO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
XX (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WM;
XX WPI; 2000-672631/65.
DR N-PSDB; AAC64619.
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 135; 139pp; English.
XX
XX The present invention describes a recombinant agfa gene (1) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEPI7/YAP) nucleation depended
CC assembly system of strains of *Salmonella*, *Escherichia coli* and
CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
CC *Enterobacteriaceae* host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
XX Sequence 151 AA;
SQ
Query Match 67.1%; Score 523; DB 3; Length 151;
Best Local Similarity 68.9%; Pred. No. 1.2e-42;
Matches 104; Conservative 20; Mismatches 42; Indels 0; Gaps 0;
QY 1 MLLKVAFAIAIVSGSALAGVVPQWGGGNNHGGNSSGPDYDQLVTRVTHEMAHQ 60
DB 1 MLLKVAFAIAIVFSSSALAGVVPQYGGGNNHGGNSSGPNSEINITYGGGNSALAQ 60
QY 61 SPARKSETTITGSGYNGADVGQADNSTITLTONGFRRNATIDQNAKNSDITVQYGG 120
DB 61 TDARNSDLTITGHHGGNGADVGQSDSSIDLTOGFNGSATLIDQNGKNSMTVQYFGG 120
QY 121 NNAALVNOTASDSSVYVROVFGNNATANQY 151
DB 121 GNGAAVDOTASNSVYVTOVFGNNATANQY 151
RESULT 15
ABR82651
ID ABR82651 standard; protein; 151 AA.
XX ABR82651;
XX
AC 04-DEC-2003 (first entry)
XX
DT
XX
DE E. coli CsgA subunit 15 kDa protein.
XX

KW Plasma protein; immune response; antibacterial; vaccine; gene therapy.
XX
XX *Escherichia coli*.
OS
XX
XX WO2003064446-A2.
PN
XX
XX 07-AUG-2003.
PD
XX
XX 30-JAN-2003; 2003WO-EP000943.
PF
XX
XX 31-JAN-2002; 2002GB-00002275.
PR
XX
XX (HANS-) HANSA MEDICAL RES AB.
PA
XX
XX Bjoerck L, Olsen A, Wikstroem M, Herwald H;
PI WPI; 2003-646136/61.
DR N-PSDB; ACR36153.
XX
XX New isolated peptide capable of binding a mammalian plasma protein,
PT useful in the manufacture of a medicament for the prevention and/or
PT treatment of a bacterial infection, such as *Escherichia coli*, *Salmonella*
PT or *Shigella* infections.
XX
PS Disclosure; Page 41-42; 42pp; English.
XX
XX The invention relates to an isolated peptide capable of binding a
XX mammalian plasma protein or of generating an immune response in a mammal
XX selected from sequences shown in ABR82647, ABR82648-49. The peptide or
XX antibody is useful for treating a bacterial infection in a human or
XX animal or in the manufacture of a medicament for the prophylactic
XX treatment of a bacterial infection, such as *Escherichia coli*, *Salmonella*
XX or *Shigella* infection. The peptide that is immobilized on a solid support
XX is also useful as a reagent for determining the ability of a plasma
XX protein to bind to bacteria. The present sequence represents an *E. coli*
XX 15 kDa protein
SQ
XX Sequence 151 AA;
Query Match 66.4%; Score 518; DB 7; Length 151;
Best Local Similarity 68.2%; Pred. No. 3.8e-42;
Matches 103; Conservative 20; Mismatches 28; Indels 0; Gaps 0;
QY 1 MLLKVAFAIAIVSGSALAGVVPQWGGGNNHGGNSSGPDYDQLVTRVTHEMAHQ 60
DB 1 MLLKVAFAIAIVFSSSALAGVVPQYGGGNNHGGNSSGPNSEINITYGGGNSALAQ 60
QY 61 SPARKSETTITGSGYNGADVGQADNSTITLTONGFRRNATIDQNAKNSDITVQYGG 120
DB 61 TDARNSDLTITGHHGGNGADVGQSDSSIDLTOGFNGSATLIDQNGKNSMTVQYFGG 120
QY 121 NNAALVNOTASDSSVYVROVFGNNATANQY 151
DB 121 GNGAAVDOTASNSVYVTOVFGNNATANQY 151
Search completed: March 11, 2004, 18:33:38
Job time : 46.9 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Comugen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:25:54 ; Search time 12.4 Seconds
(without alignments)
628,671 Million cell updates/sec

Title: US-09-543-407-18

Perfect score: 780
Sequence: 1 MTLKVAFAAIVSGSALA.....DSSVMRYQFGNNATANY 151Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 200000000Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/prodata/2/iaa/58_COMB.pep:*
2: /cgn2_6/prodata/2/iaa/58_COMB.pep:*
3: /cgn2_6/prodata/2/iaa/58_COMB.pep:*
4: /cgn2_6/prodata/2/iaa/58_COMB.pep:*
5: /cgn2_6/prodata/2/iaa/pctus_COMB.pep:*
6: /cgn2_6/prodata/2/iaa/backfillset.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	686	87.9	151	1 US-08-233-788A-59	Sequence 59, App1
2	506	64.9	120	1 US-08-233-788A-57	Sequence 57, App1
3	92	11.8	975	4 US-09-328-352-4764	Sequence 4764, Ap
4	90	11.5	906	1 US-08-254-573-2	Sequence 2, App1
5	90	11.5	906	1 US-08-687-379-2	Sequence 2, App1
6	90	11.5	906	1 US-08-687-379-4	Sequence 4, App1
7	90	11.5	906	4 US-08-172-332-1	Sequence 1, App1
8	90	11.5	906	4 US-08-216-326-2	Sequence 2, App1
9	87.5	11.2	738	3 US-08-664-038A-3	Sequence 3, App1
10	84	10.8	892	4 US-09-336-447A-5	Sequence 5, App1
11	84	10.8	907	1 US-07-718-575-2	Sequence 2, App1
12	84	10.8	907	1 US-08-481-206-2	Sequence 2, App1
13	84	10.8	907	2 US-08-486-269A-2	Sequence 2, App1
14	84	10.8	943	4 US-09-056-556-204	Sequence 204, App
15	84	10.8	943	4 US-09-072-596-199	Sequence 199, App
16	84	10.8	943	4 US-09-477-135A-131	Sequence 131, App
17	84	10.8	943	4 US-09-072-967-204	Sequence 204, App
18	84	10.8	1415	4 US-09-252-991A-26438	Sequence 26438, A
19	83.5	10.7	415	4 US-09-025-769B-280	Sequence 280, App
20	83.5	10.7	873	4 US-09-336-447A-13	Sequence 13, App1
21	83	10.6	518	3 US-09-043-123-2	Sequence 2, App1
22	83	10.6	1912	1 US-08-409-995-4	Sequence 4, App1
23	83	10.6	1912	3 US-08-685-467-4	Sequence 4, App1
24	83	10.6	2353	3 US-09-377-155-33	Sequence 33, App1
25	83	10.6	2353	3 US-08-913-942-4	Sequence 4, App1
26	83	10.6	2353	4 US-09-669-974-33	Sequence 33, App1
27	83	10.6	2353	4 US-09-797-862-33	Sequence 33, App1

28	83	10.6	2354	4 US-09-268-347-47	Sequence 47, App1
29	83	10.6	2411	4 US-09-268-347-36	Sequence 36, App1
30	82	10.5	878	4 US-09-540-236-3401	Sequence 3401, Ap
31	81.5	10.4	211	1 US-08-276-852-34	Sequence 34, App1
32	81.5	10.4	211	1 US-08-133-011-16	Sequence 16, App1
33	81.5	10.4	211	1 US-08-322-730A-16	Sequence 16, App1
34	81.5	10.4	211	1 US-08-387-874-16	Sequence 16, App1
35	81.5	10.4	211	1 US-08-899-575-34	Sequence 34, App1
36	81.5	10.4	211	1 US-08-899-575-34	Sequence 34, App1
37	81.5	10.4	211	1 US-08-383-619-16	Sequence 16, App1
38	81.5	10.4	211	3 US-08-907-739-16	Sequence 16, App1
39	81.5	10.4	211	4 US-09-729-597-16	Sequence 16, App1
40	81.5	10.4	211	5 PCT-US93-08364-16	Sequence 16, App1
41	81.5	10.4	211	5 PCT-US95-08743-34	Sequence 34, App1
42	81.5	10.4	238	4 US-09-495-880A-42	Sequence 42, App1
43	81.5	10.4	266	4 US-09-495-880A-26	Sequence 26, App1
44	81.5	10.4	293	3 US-08-438-745-4	Sequence 4, App1
45	81.5	10.4	293	3 US-08-438-745-6	Sequence 4, App1

ALIGNMENTS

RESULT 1
US-08-233-788A-59
; Sequence 59, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, Karen S.
; APPLICANT: Clouthier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; TITLE OF INVENTION: OF SALMONELLA
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,788A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 682-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEDANBERY
; INFORMATION FOR SEQ ID NO: 59:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 151 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-233-788A-59
Query Match 87.9%; Score 686; DB 1; Length 151;
Best Local Similarity 90.1%; Pred. No. 2, 2e-62;
Matches 136; Conservative 3; Mismatches 12; Indels 0; Gaps 0;
1 MTLKVAFAAIVSGSALAIVPGWGGGNNNGGNSGPRYDQLVTRVTHMAHAQ 60
|||||

Db 1 MMLKVAFAAIVAGSSALAGVVPQWGGGNGGNSGSPDSTLIYQGSANALALQ 60
 QY 61 SPARKSETTITGSGNGADVGAGADNSTITLQNGFRNNATIDQWAKNSITTVGYGG 120
 Db 61 SPARKSETTITGSGNGADVGAGADNSTITLQNGFRNNATIDQWAKNSITTVGYGG 120
 QY 121 NNAALVNOTASDSVVRQVGFQNNATANOY 151
 Db 121 NNPALVNOTASDSVVRQVGFQNNATANOY 151

RESULT 2

US-08-233-788A-57
 ; Sequence 57, Application US/08233788A
 ; Patent No. 5635617
 ; GENERAL INFORMATION:

APPLICANT: Doran, James L.
 APPLICANT: Kay, William W.
 APPLICANT: Collinson, Karen S.
 APPLICANT: Clouthier, Sharon C.
 TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
 OF SEQUENCES OF SALMONELLA
 NUMBER OF SEQUENCES: 61
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Seed and Berry
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington
 COUNTRY: U.S.A.
 ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/233,788A
 FILING DATE: 26-APR-1994
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: King, Joshua
 REGISTRATION NUMBER: 35,570
 REFERENCE/DOCKET NUMBER: 920043.403C2
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 622-4900
 TELEFAX: (206) 682-6031
 TELEX: 3723836 SEDANBERY
 INFORMATION FOR SEQ ID NO: 57:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 120 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein

US-08-233-788A-57

Query Match 64.9%; Score 506; DB 1; length 120;
 Best Local Similarity .87.5%; Pred. No. 3.4e-44;

Matches 98; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 22 VVPWGGGNGGNGGNSGSPDYDQVTRVTHEMAHALQSDARKSETTITGSGNGADV 81
 Db 1 VVPWGGGNGGNGGNSGSPDSTLIYQGSANALALQSDARKSETTITGSGNGADV 60
 QY 82 GQAGDNSTIELTQNGFRNNATIDQWAKNSITTVGYGGNNAALVNOTASDS 133
 Db 61 GQAGDNSTIELTQNGFRNNATIDQWAKNSITTVGYGGNNAALVNOTASDS 112

RESULT 3

US-09-328-352-4764
 ; Sequence 47/64, Application US/09328352
 ; Patent No. 6562958
 ; GENERAL INFORMATION:

APPLICANT: Gary L. Bretton et al.
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
 ; FILE REFERENCE: GTC99-03PA
 ; CURRENT APPLICATION NUMBER: US/09/328,352
 ; CURRENT FILING DATE: 1999-06-04
 ; NUMBER OF SEQ ID NOS: 8252
 ; SEQ ID NO 4764
 ; LENGTH: 975
 ; TYPE: PRT
 ; ORGANISM: Acinetobacter baumannii
 US-09-328-352-4764

Query Match 11.8%; Score 92; DB 4; length 975;
 Best Local Similarity 23.8%; Pred. No. 0.75;

Matches 35; Conservative 24; Mismatches 52; Indels 36; Gaps 7;

QY 15 SGALAGVVPQWGGGNGGNGGNSGSPDYDQVTRVTHEM-----AHALQSDA 63
 Db 300 AGNGIA-----SGNGEHYIGINGNGDDVD--ITAPITGVINISGNSFTLIGNSSSSSV 351
 QY 64 RKSSETTITGSGNGADVGAGADNSTITLQNGF-----RNNATIDQWAKNS 111
 Db 352 NTAPTTSTVNDNDITDNGSGGTGSGGNSGSDGLNGASGNGEHYIGINGNGDDV 411
 QY 112 DIT-----VGYGGNNAALVNOTASDS 133
 Db 412 DITPTTGTIFNFSGNSFSLIGNSSSSS 438

RESULT 4

US-08-254-573-2
 ; Sequence 2, Application US/08254573
 ; Patent No. 5610032
 ; GENERAL INFORMATION:

APPLICANT: KAMBOJ, Rajender
 APPLICANT: ELIOTT, Candace
 APPLICANT: NUTT, Stephen
 TITLE OF INVENTION: AMPA-BINDING HUMAN GLUR1 RECEPTORS
 NUMBER OF SEQUENCES: 5
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Foley & Lardner
 STREET: 1800 Diagonal Road, Suite 500
 CITY: Alexandria
 STATE: Virginia
 COUNTRY: USA
 ZIP: 22313-0299

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/254,573
 FILING DATE: 06-JUN-1994
 CLASSIFICATION: 435
 PRIOR APPLICATION NUMBER:
 APPLICATION NUMBER: US 07/896,611
 FILING DATE: 10-JUN-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: BENT, Stephen A.
 REGISTRATION NUMBER: 29,768
 REFERENCE/DOCKET NUMBER: 16777/179 ALLE
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 836-9300
 TELEFAX: (703) 683-4109
 TELEX: 899149

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
 LENGTH: 906 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein

US-08-254-573-2

Query Match 11.5%; Score 90; DB 1; Length 906;
Best Local Similarity 22.6%; Pred. No. 1.1;
Matches 45; Conservative 26; Mismatches 56; Indels 72; Gaps 10;

QY 6 VAAFAIIVSGSALAGVFPQWGGGNNHNGGNSGPDYDOL-----VTRVYTH-----M 55
DB 250 VTGFQLVNVTDTITPAKIMQW-----KNSDADHTRVDMKRPKYSALTVDGVKM 300
QY 56 AAHALSDARKSETTITQSGYGNAGADY-----GGGAD-----NSTIELIQ 94
DB 301 AEAFOSLRQ--RIDISRRGNAGDCLANPAVWGGIDIQRALQOVREGLTGNVQFNE 357
QY 95 NGFRNNAT-----IDOWNAKNSDI--TVGYGGNNALVNOT-----ASD 132
DB 358 KGRRTVYTLHVIEMKGDSTRKIGYWNEDDKFVPAATDAAGDGNSSVQRTYIVTTILBD 417
QY 133 SSVWVQVQFGNNATANQY 151
DB 418 PYVMLKK-----NANQF 429

RESULT 5

US-08-687-379-2
; Sequence 2, Application US/08687379
; Patent No. 5756697

GENERAL INFORMATION:

APPLICANT: Hoefer, Thomas
APPLICANT: Ullrich, Andreas
APPLICANT: Bach, Alfred
APPLICANT: Steier, Sylvia
APPLICANT: Lemaire, Hans-Georg
TITLE OF INVENTION: Subunits of Glutamate Receptors, Their
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESS: Keil & Weinkauff
STREET: 1101 Connecticut Avenue
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20036

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette-3.5 inch, 720 Kb storage
COMPUTER: IBM AT-compatible, 80286 processor
OPERATING SYSTEM: MS-DOS version 6.0
SOFTWARE: Wordperfect version 5.1
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/687,379
FILING DATE: 05-AUG-1996
CLASSIFICATION: 435
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 906 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-687-379-2

Query Match 11.5%; Score 90; DB 1; Length 906;
Best Local Similarity 22.6%; Pred. No. 1.1;
Matches 45; Conservative 26; Mismatches 56; Indels 72; Gaps 10;

QY 6 VAAFAIIVSGSALAGVFPQWGGGNNHNGGNSGPDYDOL-----VTRVYTH-----M 55
DB 250 VTGFQLVNVTDTITPAKIMQW-----KNSDADHTRVDMKRPKYSALTVDGVKM 300
QY 56 AAHALSDARKSETTITQSGYGNAGADY-----GGGAD-----NSTIELIQ 94
DB 301 AEAFOSLRQ--RIDISRRGNAGDCLANPAVWGGIDIQRALQOVREGLTGNVQFNE 357
QY 95 NGFRNNAT-----IDOWNAKNSDI--TVGYGGNNALVNOT-----ASD 132

DB 358 KGRRTVYTLHVIEMKGDSTRKIGYWNEDDKFVPAATDAAGDGNSSVQRTYIVTTILBD 417

QY 133 SSVWVQVQFGNNATANQY 151
DB 418 PYVMLKK-----NANQF 429

RESULT 6

US-08-687-379-4
; Sequence 4, Application US/08687379
; Patent No. 5756697

GENERAL INFORMATION:

APPLICANT: Hoefer, Thomas
APPLICANT: Ullrich, Andreas
APPLICANT: Bach, Alfred
APPLICANT: Steier, Sylvia
APPLICANT: Lemaire, Hans-Georg
TITLE OF INVENTION: Subunits of Glutamate Receptors, Their
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESS: Keil & Weinkauff
STREET: 1101 Connecticut Avenue
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20036

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette-3.5 inch, 720 Kb storage
COMPUTER: IBM AT-compatible, 80286 processor
OPERATING SYSTEM: MS-DOS version 6.0
SOFTWARE: Wordperfect version 5.1
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/687,379
FILING DATE: 05-AUG-1996
CLASSIFICATION: 435
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 906 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-687-379-4

Query Match 11.5%; Score 90; DB 1; Length 906;
Best Local Similarity 22.6%; Pred. No. 1.1;
Matches 45; Conservative 26; Mismatches 56; Indels 72; Gaps 10;

QY 6 VAAFAIIVSGSALAGVFPQWGGGNNHNGGNSGPDYDOL-----VTRVYTH-----M 55
DB 250 VTGFQLVNVTDTITPAKIMQW-----KNSDADHTRVDMKRPKYSALTVDGVKM 300
QY 56 AAHALSDARKSETTITQSGYGNAGADY-----GGGAD-----NSTIELIQ 94
DB 301 AEAFOSLRQ--RIDISRRGNAGDCLANPAVWGGIDIQRALQOVREGLTGNVQFNE 357
QY 95 NGFRNNAT-----IDOWNAKNSDI--TVGYGGNNALVNOT-----ASD 132
DB 358 KGRRTVYTLHVIEMKGDSTRKIGYWNEDDKFVPAATDAAGDGNSSVQRTYIVTTILBD 417
QY 133 SSVWVQVQFGNNATANQY 151
DB 418 PYVMLKK-----NANQF 429

RESULT 7

US-08-172-332-1
; Sequence 1, Application US/08172332
; Patent No. 6313279

GENERAL INFORMATION:

APPLICANT: Burnette, J. Paul
APPLICANT: Mayne, Nancy G

```

;
; APPLICANT: Snyder, Yvonne M
; APPLICANT: Sharp, Robert L
; TITLE OF INVENTION: HUMAN GLUTAMATE RECEPTOR AND RELATED DNA
; TITLE OF INVENTION: COMPOUNDS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESS: Lilly Patent Division/JPL
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/172,332
; FILING DATE: 22-DEC-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/879,688
; FILING DATE: May 1, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Leeds, James P.
; REGISTRATION NUMBER: 35241
; REFERENCE/DOCKET NUMBER: X-8342
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-1667
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 906 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-172-332-1

Query Match      11.5%; Score 90; DB 4; Length 906;
Best Local Similarity 22.6%; Pred. No. 1.1;
Matches 45; Conservative 26; Mismatches 56; Indels 72; Gaps 10;

QY 6 VAAFAAIIVSSGALAGVVPQWGGGNGGNSGSPDYDQ-----VTRVTHHE---M 55
DB 250 VTGFQLVNVTDTIPAKIMQW-----KNSDARDHTRVDWKRPKYTSALTVDGKVM 300
QY 56 AAHALQSDARKSETTITOSGYNGADV-----GQGAD-----NSTIELTQ 94
DB 301 ABAFQSLRRQ---RIDISRRGNAGDCLANPAVPWGQSIDIQRALQOVREGLTGNVQFNE 357
QY 95 NGFRNNAT-----IDWNNAKNSDI---TVGYGNNALVNQT-----ASD 132
DB 358 KGRRTNVTTLHVLVEMKHGDIRKIGYWNEDDKFVPATDAQAAGDSSVQNRITYITLTIED 417
QY 133 SSVWVROVGFNNATANQY 151
DB 418 PYWMLKK-----NANQF 429

RESULT 8
US-08-216-326-2
; Sequence 2, Application US/08216326
; Patent No. 6406868
; GENERAL INFORMATION:
; APPLICANT: KAMBOU, Rajender
; APPLICANT: ELIOTT, Stephen
; APPLICANT: NUTT, Stephen
; TITLE OF INVENTION: AMPA-BINDING HUMAN GRI1 RECEPTORS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street N.W., Suite 500
; CITY: Washington, D.C.

```

```

;
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/216,326
; FILING DATE: 23-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/896,611
; FILING DATE: 10-JUN-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 16777/229/ALLE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 906 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-216-326-2

Query Match      11.5%; Score 90; DB 4; Length 906;
Best Local Similarity 22.6%; Pred. No. 1.1;
Matches 45; Conservative 26; Mismatches 56; Indels 72; Gaps 10;

QY 6 VAAFAAIIVSSGALAGVVPQWGGGNGGNSGSPDYDQ-----VTRVTHHE---M 55
DB 250 VTGFQLVNVTDTIPAKIMQW-----KNSDARDHTRVDWKRPKYTSALTVDGKVM 300
QY 56 AAHALQSDARKSETTITOSGYNGADV-----GQGAD-----NSTIELTQ 94
DB 301 ABAFQSLRRQ---RIDISRRGNAGDCLANPAVPWGQSIDIQRALQOVREGLTGNVQFNE 357
QY 95 NGFRNNAT-----IDWNNAKNSDI---TVGYGNNALVNQT-----ASD 132
DB 358 KGRRTNVTTLHVLVEMKHGDIRKIGYWNEDDKFVPATDAQAAGDSSVQNRITYITLTIED 417
QY 133 SSVWVROVGFNNATANQY 151
DB 418 PYWMLKK-----NANQF 429

RESULT 9
US-08-864-038A-3
; Sequence 3, Application US/08864038A
; Patent No. 6001592
; GENERAL INFORMATION:
; APPLICANT: KUNIO NAKASHIMA et al.
; TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE cDNA, VECTOR
; TITLE OF INVENTION: CONTAINING SAID cDNA, HOST CELLS TRANSFORMED WITH SAID
; TITLE OF INVENTION: VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING
; TITLE OF INVENTION: SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBODY
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: 812-5 Hirano
; STREET: Ieshinden
; CITY: Tsu-city
; STATE: Mie-prefecture
; COUNTRY: JAPAN
; ZIP: 514-01
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
; COMPUTER: IBM compatible
; OPERATING SYSTEM: Microsoft windows 95

```

```

1      SOFTWARE: word Perfect 6.1
2      CURRENT APPLICATION DATA:
3      APPLICATION NUMBER: US/08/864,038A
4      FILING DATE: May 28, 1997
5      PRIOR APPLICATION DATA:
6      APPLICATION NUMBER: JP 8-184459
7      FILING DATE: 15-July-1996
8      ATTORNEY/AGENT INFORMATION:
9      NAME: C. Bruce Hamburg
10     REGISTRATION NUMBER: 22,389
11     REFERENCE/DOCKET NUMBER: F-5610
12     TELECOMMUNICATION INFORMATION:
13     TELEPHONE: (212)986-2340
14     TELEFAX: (212)953-7733
15     INFORMATION FOR SEQ ID NO: 3:
16     SEQUENCE CHARACTERISTICS:
17     LENGTH: 738
18     TYPE: amino acid
19     TOPOLOGY: linear
20     MOLECULE TYPE: protein
21     ORIGINAL SOURCE:
22     ORGANISM: Pinctada fucata
23     CELL TYPE: mantle epithelial cell
24     FEATURE:
25     NAME/KEY: peptide
26     LOCATION: from 1 to 738
27     IDENTIFICATION METHOD: E (by experiment)
28     US-08-864-038A-3

```

QY 142 FGNNATANOY 151
 Db 425 -----NANOF 429

RESULT 12

US-08-481-206-2
 / Sequence 2, Application US/08481206
 / Patent No. 5739291
 / GENERAL INFORMATION:
 / APPLICANT: Heinemann Ph.D., Stephen F.
 / APPLICANT: Boulter Ph.D., James R.
 / APPLICANT: Hollmann Ph.D., Michael NMN
 / APPLICANT: Bettler Ph.D., Bernhard NMN
 / APPLICANT: Jensen Ph.D., Jan E.
 / TITLE OF INVENTION: GLUTAMATE RECEPTOR COMPOSITIONS AND
 / METHODS
 / NUMBER OF SEQUENCES: 14
 / CORRESPONDENCE ADDRESS:
 / ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
 / STREET: 444 So. Flower St., Suite 2000
 / CITY: Los Angeles
 / STATE: California
 / COUNTRY: United States
 / ZIP: 90071-2921
 / COMPUTER READABLE FORM:
 / MEDIUM TYPE: Floppy disk
 / COMPUTER: IBM PC compatible
 / OPERATING SYSTEM: PC-DOS/MS-DOS
 / SOFTWARE: Patent in Release #1.0, Version #1.25
 / CURRENT APPLICATION DATA:
 / APPLICATION NUMBER: US/08/481,206
 / FILING DATE:
 / CLASSIFICATION: 530
 / PRIOR APPLICATION DATA:
 / APPLICATION NUMBER: US/08/013,767
 / FILING DATE:
 / ATTORNEY/AGENT INFORMATION:
 / NAME: Reiter Ph.D., Stephen E.
 / REGISTRATION NUMBER: 31192
 / REFERENCE/DOCKET NUMBER: P31 8962
 / TELECOMMUNICATION INFORMATION:
 / TELEPHONE: (619) 535-9001
 / TELEFAX: (619) 535-8949
 / TELEX: 9103330318
 / INFORMATION FOR SEQ ID NO: 2:
 / SEQUENCE CHARACTERISTICS:
 / LENGTH: 907 amino acids
 / TYPE: amino acid
 / TOPOLOGY: linear
 / MOLECULE TYPE: protein
 / US-08-481-206-2

Query Match 10.8%; Score 84; DB 1; Length 907;
 Best Local Similarity 22.6%; Pred. No. 4.4;
 Matches 43; Conservative 24; Mismatches 69; Indels 54; Gaps 8;

QY 6 VAAFAIVVSSALAGVVPQWGGGNNNGG-NSSGPDYDQVTRVTHVEMHALQSDAR 64
 Db 250 VTGQLVVYDTITPARIMQWRTSDSRDHTRVDMKPKYTSALTIDGVKVAEAFQSLRR 309
 QY 65 KSETTTTQSGYGAGADV-----GQAD-----NSTIELTONGFRNNAT- 102
 Db 310 Q---RIDISRKAGAGCLANPAVPMQGGIDIDRALQVREFGLTGNVQFNEKGRRTNYTL 366
 QY 103 -----IDQNAKNSDI---TVGQYGNNALVNOT-----ASDSSVWVRQVG 141
 Db 367 HVIEMKHDGIRKIGYWNEDDKFVPATDAQAGSDNSVQNTIYVTTILDEPYMLK-- 424
 QY 142 FGNNATANOY 151
 Db 425 -----NANOF 429

RESULT 13

US-08-486-269A-2
 / Sequence 2, Application US/08486269A
 / Patent No. 5945509
 / GENERAL INFORMATION:
 / APPLICANT: Heinemann, Stephen F.
 / APPLICANT: Boulter, James R.
 / APPLICANT: Hollmann, Michael
 / APPLICANT: Bettler, Bernhard
 / APPLICANT: Jensen, Jan E.
 / TITLE OF INVENTION: GLUTAMATE RECEPTOR COMPOSITIONS
 / AND METHODS
 / NUMBER OF SEQUENCES: 14
 / CORRESPONDENCE ADDRESS:
 / ADDRESSEE: Gray Cary Ware & Freidenrich LLP
 / STREET: 4365 Executive Drive, Suite 1600
 / CITY: San Diego
 / STATE: CA
 / COUNTRY: USA
 / ZIP: 92121
 / COMPUTER READABLE FORM:
 / MEDIUM TYPE: Diskette
 / COMPUTER: IBM Compatible
 / OPERATING SYSTEM: DOS
 / SOFTWARE: FastSeq for Windows DEMONSTRATION Version 2.0D
 / CURRENT APPLICATION DATA:
 / APPLICATION NUMBER: US/08/486,269A
 / FILING DATE: 06-JUN-1995
 / CLASSIFICATION: 530
 / PRIOR APPLICATION DATA:
 / APPLICATION NUMBER: 08/013,767
 / FILING DATE: 04-FEB-1993
 / APPLICATION NUMBER: 07/718,575
 / FILING DATE: 21-JUN-1991
 / APPLICATION NUMBER: PCT/US90/06153
 / FILING DATE: 25-OCT-1990
 / APPLICATION NUMBER: 07/428,116
 / FILING DATE: 27-OCT-1989
 / ATTORNEY/AGENT INFORMATION:
 / NAME: Reiter, Stephen E.
 / REGISTRATION NUMBER: 31,192
 / REFERENCE/DOCKET NUMBER: P41 9986
 / TELECOMMUNICATION INFORMATION:
 / TELEPHONE: 619-677-1409
 / TELEFAX: 619-677-1465
 / TELEX:
 / INFORMATION FOR SEQ ID NO: 2:
 / SEQUENCE CHARACTERISTICS:
 / LENGTH: 907 amino acids
 / TYPE: amino acid
 / STRANDEDNESS: single
 / TOPOLOGY: linear
 / MOLECULE TYPE: protein
 / US-08-486-269A-2

Query Match 10.8%; Score 84; DB 2; Length 907;
 Best Local Similarity 22.6%; Pred. No. 4.4;
 Matches 43; Conservative 24; Mismatches 69; Indels 54; Gaps 8;

QY 6 VAAFAIVVSSALAGVVPQWGGGNNNGG-NSSGPDYDQVTRVTHVEMHALQSDAR 64
 Db 250 VTGQLVVYDTITPARIMQWRTSDSRDHTRVDMKPKYTSALTIDGVKVAEAFQSLRR 309
 QY 65 KSETTTTQSGYGAGADV-----GQAD-----NSTIELTONGFRNNAT- 102
 Db 310 Q---RIDISRKAGAGCLANPAVPMQGGIDIDRALQVREFGLTGNVQFNEKGRRTNYTL 366
 QY 103 -----IDQNAKNSDI---TVGQYGNNALVNOT-----ASDSSVWVRQVG 141
 Db 367 HVIEMKHDGIRKIGYWNEDDKFVPATDAQAGSDNSVQNTIYVTTILDEPYMLK-- 424

QY 142 FGNMATAOY 151
DB 425 -----NANOF 429

RESULT 14

US-09-056-556-204

/ Sequence 204, Application US/09056556
/ Patent No. 6350456
/ GENERAL INFORMATION:

APPLICANT: Reed, Steven G.

APPLICANT: Skeiky, Yasir A.W.

APPLICANT: Dillon, Davin C.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND

NUMBER OF SEQUENCES: 241

CORRESPONDENCE ADDRESS:

ADDRESS: SEED and BERRY LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/056,556

FILING DATE: 07-APR-1998

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Maki, David J.

REGISTRATION NUMBER: 31,392

REFERENCE/DOCKET NUMBER: 210121.457

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 204:

SEQUENCE CHARACTERISTICS:

LENGTH: 943 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

US-09-056-556-204

Query Match 10.8%; Score 84; DB 4; Length 943;

Best Local Similarity 25.2%; Pred. No. 4.7;

Matches 31; Conservative 9; Mismatches 41; Indels 42; Gaps 4;

QY 15 SCSALAGVVPWGGGNNHNGGNSGPDYDQVTRVVTHEMAHALQSDARKSETTITGSG 74
DB 184 SGTGNVGI-----GNSGTGMWIGIGNSGNSYN-----TG 211

QY 75 YENGADVGGADNSTIELTONGFRNNATIDOMNAKNSD---ITVGOY-----GGNNA 124
DB 212 FGNSGDANTGFNNSGIANTGVNAGNYNTGSPNSNTGFRMGQYNTGYLNSGNYNTG 271

QY 125 LVN 127
DB 272 LAN 274

RESULT 15

US-09-072-596-199

/ Sequence 199, Application US/09072596
/ Patent No. 6458366
/ GENERAL INFORMATION:

APPLICANT: Reed, Steven G.

APPLICANT: Skeiky, Yasir A.W.

APPLICANT: Dillon, Davin C.

APPLICANT: Campos-Neto, Antonia

APPLICANT: Houghton, Raymond

APPLICANT: Vedvick, Thomas S.
APPLICANT: Iwardzik, Daniel R.
APPLICANT: Lodes, Michael J.
APPLICANT: Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
NUMBER OF SEQUENCES: 350
CORRESPONDENCE ADDRESS:
ADDRESS: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/072,596
FILING DATE: 05-MAY-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C9
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 199:
SEQUENCE CHARACTERISTICS:
LENGTH: 943 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear

US-09-072-596-199

Query Match 10.8%; Score 84; DB 4; Length 943;

Best Local Similarity 25.2%; Pred. No. 4.7;

Matches 31; Conservative 9; Mismatches 41; Indels 42; Gaps 4;

QY 15 SCSALAGVVPWGGGNNHNGGNSGPDYDQVTRVVTHEMAHALQSDARKSETTITGSG 74
DB 184 SGTGNVGI-----GNSGTGMWIGIGNSGNSYN-----TG 211

QY 75 YENGADVGGADNSTIELTONGFRNNATIDOMNAKNSD---ITVGOY-----GGNNA 124
DB 212 FGNSGDANTGFNNSGIANTGVNAGNYNTGSPNSNTGFRMGQYNTGYLNSGNYNTG 271

QY 125 LVN 127
DB 272 LAN 274

Search completed: March 11, 2004, 18:44:51
Job time: 13.4 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:40:56 ; Search time 24.6 Seconds
(without alignments)
1296.103 Million cell updates/sec

Title: US-09-543-407-18

Perfect score: 780
Sequence: 1 MRLKVAFAAIWGSALA.....DSSVMROYFGNNATANY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 809742 seque, 21153259 residues

Total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	108.5	13.9	445	US-10-369-493-20638	Sequence 20638, A
2	91	11.7	597	US-09-793-306-146	Sequence 146, App
3	90	11.5	906	US-10-251-661-2	Sequence 2, Appl1
4	85.5	11.0	278	US-09-810-264-28	Sequence 28, Appl1
5	84.5	10.8	408	US-10-369-493-12833	Sequence 12833, A
6	84	10.8	892	US-09-952-267-5	Sequence 5, App
7	84	10.8	943	US-09-965-634-131	Sequence 131, App
8	84	10.8	943	US-09-997-182-131	Sequence 131, App
9	84	10.8	943	US-09-997-181-131	Sequence 131, App
10	84	10.8	943	US-10-193-002-159	Sequence 159, App
11	84	10.7	943	US-10-084-843-204	Sequence 204, App
12	83.5	10.8	65	US-09-996-194-16	Sequence 16, Appl
13	83.5	10.7	873	US-09-952-267-13	Sequence 13, Appl
14	83	10.6	182	US-10-238-075-1549	Sequence 1549, Ap
15	83	10.6	354	US-09-820-843A-21	Sequence 21, Appl

16	83	10.6	518	9	US-09-976-297-2	Sequence 2, Appl1
17	83	10.6	720	9	US-09-801-368-176	Sequence 176, App
18	83	10.6	253	9	US-09-797-862-33	Sequence 33, Appl
19	82.5	10.6	434	15	US-10-045-674-594	Sequence 594, App
20	81.5	10.4	211	14	US-10-016-986-34	Sequence 34, Appl
21	81.5	10.4	211	15	US-10-273-973-16	Sequence 16, Appl
22	81.5	10.4	219	9	US-09-809-517A-24	Sequence 24, Appl
23	81.5	10.4	219	9	US-09-809-517A-27	Sequence 27, Appl
24	81.5	10.4	287	14	US-10-033-399B-12	Sequence 12, Appl
25	81.5	10.4	319	9	US-09-884-767A-215	Sequence 215, App
26	81.5	10.4	424	9	US-09-884-767A-213	Sequence 213, App
27	81.5	10.4	424	15	US-10-045-674-591	Sequence 591, App
28	81.5	10.4	432	9	US-09-809-517A-25	Sequence 25, Appl
29	81.5	10.4	434	9	US-09-809-517A-23	Sequence 23, Appl
30	81.5	10.4	434	9	US-09-809-517A-26	Sequence 26, Appl
31	81.5	10.4	484	11	US-09-896-095-272	Sequence 272, App
32	81.5	10.4	533	15	US-10-045-674-527	Sequence 527, App
33	80.5	10.3	550	14	US-10-156-761-11286	Sequence 11286, A
34	80	10.3	232	14	US-10-076-604-103	Sequence 103, App
35	80	10.3	233	14	US-10-076-604-101	Sequence 101, App
36	80	10.3	233	14	US-10-076-604-105	Sequence 105, App
37	80	10.3	273	15	US-10-369-493-20096	Sequence 20096, A
38	80	10.3	594	14	US-10-156-761-13173	Sequence 13173, A
39	80	10.3	939	14	US-10-156-761-12605	Sequence 12605, A
40	79.5	10.2	256	14	US-10-156-761-11972	Sequence 11972, A
41	79.5	10.2	956	13	US-10-121-032-63	Sequence 63, App
42	79.5	10.2	956	14	US-10-093-037-61	Sequence 63, Appl
43	79.5	10.2	2732	14	US-10-238-075-1119	Sequence 1119, Ap
44	79.5	10.2	2834	14	US-10-085-959-252	Sequence 252, App
45	78.5	10.1	274	14	US-10-156-761-9107	Sequence 9107, Ap

ALIGNMENTS

RESULT 1
US-10-369-493-20638
; Sequence 20638, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xiangeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 20638
; LENGTH: 445
; TYPE: PRT
; ORGANISM: Rhodospseudomonas palustris
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(445)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-20638

Query Match 13.9%; Score 108.5; DB 15; Length 445;
Best Local Similarity 27.5%; Pred. No. 0.0089;

Matches 46; Conservative 19; Mismatches 61; Indels 41; Gaps 6;

QY 7 AAFPA-----IVSGSALAGVPPQWGGG-----NHNG-----GNSGPDYDQVTRV 50
Db 19 AAFADSNVTIVNQCNDQANITQSGNGNSVGAFFNGSGFLQENGTLLG---LTL-- 73	
QY 51 VTHMAHAIQSARKESTTTTSGVGNAGDVQAGADNSTIELTQNGFRNNATIDQNNM 110	

Db 74 -----VKSGNSGNSVGRDIOGKSGAGNSAIFEGEGTGDVLEQOTGTSNCAVPSGMMWTN 129
 Qy 111 -----SIIITGYGGNNAALVNOTASDSSVMRQVG 141
 Db 130 DPGVFNKITDSSSSGKSVIODGKNVFSIKOGTGNSTSVNQIG 176

RESULT 2
 US-09-793-306-146
 ; Sequence 146, Application US/09793306
 ; Patent No. US20020098200A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Campos-Neco, Antonio
 ; APPLICANT: Skeiky, Yasir
 ; APPLICANT: Owendale, Pamela
 ; APPLICANT: Jen, Shylan
 ; APPLICANT: Lodes, Michael
 ; APPLICANT: Corixa Corporation
 ; TITLE OF INVENTION: Compounds and Methods for Diagnosis and Immunotherapy
 ; FILE REFERENCE: 014058-008740US
 ; CURRENT APPLICATION NUMBER: US/09/793,306
 ; PRIOR FILING DATE: 2001-02-26
 ; PRIOR APPLICATION NUMBER: US 60/185,037
 ; PRIOR FILING DATE: 2000-02-25
 ; PRIOR APPLICATION NUMBER: US 60/223,828
 ; PRIOR FILING DATE: 2000-08-08
 ; NUMBER OF SEQ ID NOS: 164
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 146
 ; LENGTH: 597
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence:mttc#3-His
 US-09-793-306-146

Query Match 11.7%; Score 91; DB 9; Length 597;
 Best Local Similarity 27.8%; Pred. No. 0.8;
 Matches 35; Conservative 13; Mismatches 48; Indels 30; Gaps 5;
 Qy 26 WGGGNNHNGGNSGPDYDQVTRVVTHEMAHALQSDARKSETTITGSGYGNAGADYQGA 85
 Db 358 FGNSGNNNIGFENSG-----NNNVGFNSGNNNFGGAGADITGFG 398
 Qy 86 DNSTIELTQNGFRNNAATIDQ--NNAKSDITVGYGGNNAALVNOTASDSSVMRQVG 143
 Db 399 GNAAGD--TNTGFGNAGFPNMGIGNAGNEDMGVNGSGSFNVGVGN--AGNQS-----VGFG 449
 Qy 144 NNATAN 149
 Db 450 NAGTILN 455

RESULT 3
 US-10-251-661-2
 ; Sequence 2, Application US/10251661
 ; Publication No. US2003016655A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Alberini, Cristina M.
 ; APPLICANT: Bear, Mark P.
 ; TITLE OF INVENTION: Methods and Compositions for Regulating
 ; FILE REFERENCE: 3499,1001-003
 ; CURRENT APPLICATION NUMBER: US/10/251,661
 ; PRIOR FILING DATE: 2002-09-20
 ; PRIOR APPLICATION NUMBER: 60/193,614
 ; PRIOR FILING DATE: 2000-03-31
 ; PRIOR APPLICATION NUMBER: PCT/US01/10661
 ; PRIOR FILING DATE: 2001-04-02
 ; NUMBER OF SEQ ID NOS: 12
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 2

LENGTH: 906
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-10-251-661-2

Query Match 11.5%; Score 90; DB 14; Length 906;
 Best Local Similarity 22.6%; Pred. No. 1.7;
 Matches 45; Conservative 26; Mismatches 56; Indels 72; Gaps 10;

Qy 6 VAAFAALVSGSALAGVPPQMGGNNHNGGNSGPDYDQ-----VTRVVTHE-----M 55
 Db 250 VTGFQVNVYDTPAIIQOM-----KNSDADHTRVDMKRPKTSALTVDGVKM 300
 Qy 56 AHALQSDARKSETTITGSGYGNAGADY-----GGAGD-----NSTIELQ 94
 Db 301 ABAFOSLRQ--RIDISRGNAAGDCLANPAIPKQGGIDICALLQVAEFGITGVNQFNE 357
 Qy 95 NGRNNAT-----IDMNAKNSDI--TVQYGGNNAALVNOT-----ASD 132
 Db 358 KGRRTVYTLHVIEMKHDGIRKIGYWNEDKFPVPAATDAQAGDSSVQNRFTYVTTILED 417
 Qy 133 SSVMVRQVGFNNATANY 151
 Db 418 PYVMLK-----NANQF 429

RESULT 4
 US-09-810-264-28
 ; Sequence 28, Application US/09810264
 ; Patent No. US20020076775A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Crane, Virginia C.
 ; APPLICANT: Famodu, Omolayo O.
 ; APPLICANT: Hu, Xu
 ; APPLICANT: Lu, Guhua
 ; APPLICANT: Zhang, Lingyu
 ; TITLE OF INVENTION: WRKY Transcription Factors and Methods
 ; FILE REFERENCE: 1183
 ; CURRENT APPLICATION NUMBER: US/09/810,264
 ; PRIOR FILING DATE: 2001-03-16
 ; PRIOR APPLICATION NUMBER: US 60/190,467
 ; PRIOR FILING DATE: 2000-03-17
 ; NUMBER OF SEQ ID NOS: 43
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 28
 ; LENGTH: 278
 ; TYPE: PRT
 ; ORGANISM: Triticum aestivum
 US-09-810-264-28

Query Match 11.0%; Score 85.5; DB 9; Length 278;
 Best Local Similarity 20.0%; Pred. No. 1.1;
 Matches 35; Conservative 25; Mismatches 54; Indels 61; Gaps 5;

Qy 15 SGAALAGVPPQMGGNNHNGGNSGPDYDQVTRVVTHEMAHALQSDARKSETTITGSG 74
 Db 71 NSSGCAVIAE-----DHTNGSEHSGPTPENSVTFEGDEANGAPEYKRRK-----EHG 121
 Qy 75 YNGAGADVQGA-----DNSTIELTQNGFR----- 98
 Db 122 DNEGSSGGTGAACVCPVREPRRLVQTLSDIDLDGFRMRKYQKRVKGNPNRPSYKCT 181
 Qy 99 -----NNATIDMNAKNS--DIVGYGGNNAALVNOTASDSSVMY 137
 Db 182 VGCPRVKHVERASHDRAVITTYEGRSHDVEVGAGASRALPTSSSDSSSVV 236

RESULT 5
 US-10-369-493-12833
 ; Sequence 12833, Application US/10369493
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:

```

; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 4/374
; SEQ ID NO 12833
; LENGTH: 408
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
; NAME/KEY: unsure
; LOCATION: (1)..(408)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-12833

```

```

Query Match          10.8%; Score 84.5; DB 15; Length 408;
Best Local Similarity 28.3%; Pred. No. 2.3;
Matches 30; Conservative 13; Mismatches 52; Indels 11; Gaps 2;

```

```

QY 35 GGNSSGPDYDQVTRVVTHEMAHALQSDARKSETTITGSGYGNADVGAGADNSTLTQ 94
DB 18 GSTSTTDYSGITSAVASCRAIVLSNVAPNESALDLSLQMGTTVTG-----LTT 71
QY 95 NGFRNATIDQWNAKNSDITV-----GQYGNNAALVNTGASPSV 135
DB 72 FGFTNSSFPDINIGKXDTITTEGSDVIDNGQAVMDGSGNGV 117

```

```

RESULT 6
US-09-952-267-5
; Sequence 5, Application US/09952267
; Publication No. US20030032772A1
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USP41 AND USP42 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/09/952,267
; CURRENT FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: 09/336,447
; PRIOR FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 892
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-952-267-5

```

```

Query Match          10.8%; Score 84; DB 10; Length 892;
Best Local Similarity 26.8%; Pred. No. 7;
Matches 41; Conservative 17; Mismatches 55; Indels 40; Gaps 9;

```

```

QY 28 GGGNHNNGGNS--GPDYDQVTRVVTHEMAHALQSDARKSETTIT--TOSGYNGAD 80
DB 75 GGDNEAKNGYVSGGADYNEAKNGYST--VGGSSNTAKGKSTIGGDTNDANGTYST 132
QY 81 VGG-----ADNSTI-----ELTNGFRNATIDQWNAKNSDITVQYG---G 120
DB 133 IGGGYISRAIGDSSTIGGGYINQATGKSTVAGGRN---QATGNNSTVAGGSINQATG 188

```

```

QY 121 NNALV---NOTASDSYMRQVGHGNNATTAN 149
DB 189 NNSTVAGSGSHQATGEGSF---AAGVENKANNAN 218

```

```

RESULT 7
US-09-996-634-131
; Sequence 131, Application US/09996634
; Patent No. US20020172684A1
; GENERAL INFORMATION:
; APPLICANT: Nano, Francis
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
; TITLE OF INVENTION: Immunostimulatory Peptides
; FILE REFERENCE: 61260
; CURRENT APPLICATION NUMBER: US/09/996,634
; CURRENT FILING DATE: 2001-11-28
; PRIOR APPLICATION NUMBER: 09/447,135
; PRIOR FILING DATE: 2000-01-03
; PRIOR APPLICATION NUMBER: 08/990,823
; PRIOR FILING DATE: 1997-12-15
; PRIOR APPLICATION NUMBER: US 96/10375
; PRIOR FILING DATE: 1996-06-14
; PRIOR APPLICATION NUMBER: 60/000,254
; PRIOR FILING DATE: 1995-06-15
; NUMBER OF SEQ ID NOS: 169
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 131
; LENGTH: 943
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-996-634-131

```

```

Query Match          10.8%; Score 84; DB 9; Length 943;
Best Local Similarity 25.2%; Pred. No. 7.5;
Matches 31; Conservative 9; Mismatches 41; Indels 42; Gaps 4;

```

```

QY 15 GSGALAGVVPQWGGGNNHNGGNSGPDYDQVTRVVTHEMAHALQSDARKSETTITGSG 74
DB 291 SGTGVNGI-----GNSGTGMWIGJGNSGNSYN-----TG 318
QY 75 YENGADVGAGADNSTITELTNGFRNATIDQWNAKNSD---ITVGY-----GNNNA 124
DB 319 FNSGDPANTGFPNSGIANTGVNAGVNTGTSYNGNSNTGFMQYNTGTGNSGNYNTG 378
QY 125 LVN 127
DB 379 LAN 381

```

```

RESULT 8
US-09-997-182-131
; Sequence 131, Application US/09997182
; Publication No. US20030049263A1
; GENERAL INFORMATION:
; APPLICANT: Nano, Francis
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
; TITLE OF INVENTION: Immunostimulatory Peptides
; FILE REFERENCE: 61258
; CURRENT APPLICATION NUMBER: US/09/997,182
; CURRENT FILING DATE: 2001-11-28
; PRIOR APPLICATION NUMBER: 09/447,135
; PRIOR FILING DATE: 2000-01-03
; PRIOR APPLICATION NUMBER: 08/990,823
; PRIOR FILING DATE: 1997-12-15
; PRIOR APPLICATION NUMBER: US 96/10375
; PRIOR FILING DATE: 1996-06-14
; PRIOR APPLICATION NUMBER: 60/000,254
; PRIOR FILING DATE: 1995-06-15
; NUMBER OF SEQ ID NOS: 169
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 131
; LENGTH: 943
; TYPE: PRT

```

ORGANISM: Mycobacterium tuberculosis
US-09-997-182-131

Query Match 10.8%; Score 84; DB 10; Length 943;
Best Local Similarity 25.2%; Pred. No. 7.5;
Matches 31; Conservative 9; Mismatches 41; Indels 42; Gaps 4;

QY 15 SSSALAGVVPQWGGGNNHNGGNSGPDYDQLVTRVVTTHMAHALQSDARKSETTITQSG 74
DB 291 SGTGNVGI-----GNSGTGNWGI GNSGNSYN-----TG 318
QY 75 YGNGADVQGGADNSTIELTQNGFRNNATIDQWNAKNSD---ITVGOY-----GNNNA 124
DB 319 FGNSSGDANTGFNSGIANTGVGNAGNYNTGSYNPGNSNTGFGFNMGOYNTGYLNSGNYNTG 378
QY 125 LVN 127
DB 379 LAN 381

RESULT 9

US-09-997-181-131
Sequence 131, Application US/09997181
Publication No. US20030049269A1
GENERAL INFORMATION:
APPLICANT: Nano, Francis
TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
FILE REFERENCE: 61257
CURRENT APPLICATION NUMBER: US/09/997,181
PRIOR FILING DATE: 2001-11-28
PRIOR APPLICATION NUMBER: 09/447,135
PRIOR FILING DATE: 2000-01-03
PRIOR APPLICATION NUMBER: 08/990,823
PRIOR FILING DATE: 1997-12-15
PRIOR APPLICATION NUMBER: US 96/10375
PRIOR FILING DATE: 1996-06-14
PRIOR APPLICATION NUMBER: 60/000,254
PRIOR FILING DATE: 1995-06-15
NUMBER OF SEQ ID NOS: 169
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 131
LENGTH: 943
TYPE: PRT
ORGANISM: Mycobacterium tuberculosis
US-09-997-181-131

Query Match 10.8%; Score 84; DB 10; Length 943;
Best Local Similarity 25.2%; Pred. No. 7.5;
Matches 31; Conservative 9; Mismatches 41; Indels 42; Gaps 4;

QY 15 SSSALAGVVPQWGGGNNHNGGNSGPDYDQLVTRVVTTHMAHALQSDARKSETTITQSG 74
DB 291 SGTGNVGI-----GNSGTGNWGI GNSGNSYN-----TG 318
QY 75 YGNGADVQGGADNSTIELTQNGFRNNATIDQWNAKNSD---ITVGOY-----GNNNA 124
DB 319 FGNSSGDANTGFNSGIANTGVGNAGNYNTGSYNPGNSNTGFGFNMGOYNTGYLNSGNYNTG 378
QY 125 LVN 127
DB 379 LAN 381

RESULT 10

US-10-193-002-199
Sequence 199, Application US/10193002
Publication No. US20030135026A1
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
Skelky, Yasir A.W.
Dillon, Davin C.
Campos-Neto, Antonia

Houghton, Raymond
Vedick, Thomas S.
Twardzik, Daniel R.
Lodes, Michael J.
Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
TUBERCULOSIS
NUMBER OF SEQUENCES: 350
CORRESPONDENCE ADDRESS:
ADDRESSER: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
City: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/193,002
FILING DATE: 10-Jul-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/072,596
FILING DATE: 05-MAY-1998
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121,417C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 199:
SEQUENCE CHARACTERISTICS:
LENGTH: 943 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 199:
US-10-193-002-199

Query Match 10.8%; Score 84; DB 14; Length 943;
Best Local Similarity 25.2%; Pred. No. 7.5;
Matches 31; Conservative 9; Mismatches 41; Indels 42; Gaps 4;

QY 15 SSSALAGVVPQWGGGNNHNGGNSGPDYDQLVTRVVTTHMAHALQSDARKSETTITQSG 74
DB 184 SGTGNVGI-----GNSGTGNWGI GNSGNSYN-----TG 211
QY 75 YGNGADVQGGADNSTIELTQNGFRNNATIDQWNAKNSD---ITVGOY-----GNNNA 124
DB 212 FGNSSGDANTGFNSGIANTGVGNAGNYNTGSYNPGNSNTGFGFNMGOYNTGYLNSGNYNTG 271
QY 125 LVN 127
DB 272 LAN 274

RESULT 11

US-10-084-843-204
Sequence 204, Application US/10084843
Publication No. US20030143243A1
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
Skelky, Yasir A.W.
Dillon, Davin C.
Campos-Neto, Antonio
Houghton, Raymond
Vedick, Thomas S.
Twardzik, Daniel R.
Lodes, Michael J.

```

PRIORITY APPLICATION NUMBER: 60/250,338
PRIOR FILING DATE: 2000-11-30
NUMBER OF SEQ ID NOS: 22
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 16
LENGTH: 65
TYPE: PR1
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Pfam consensus sequence
US-09-996-194-16

Query Match
Best Local Similarity 10.7%; Score 83.5; DB 9; Length 65;
Matches 21; Conservative 3; Mismatches 13; Indels 7; Gaps 1;

Cy 16 GSALAGVPPQWGGGNNHGGSSGPDYDOLVTVVTHEMAHL 59
Db 17 GRAPVGI-----NIPANITSRNHHYDQLVTVVTHEMAHL 53

RESULT 13
US-09-952-267-13
Sequence 13, Application US/09952267
Publication No. US20030032772A1
GENERAL INFORMATION:
APPLICANT: HANSEN, ERIC J.
APPLICANT: AEBI, CHRISTOPH
APPLICANT: COPE, LESLIE D.
APPLICANT: MACIVER, ISOBEL
APPLICANT: FISKE, MICHAEL J.
APPLICANT: FREDENBURG, ROSS A.
TITLE OF INVENTION: USP21 AND USP22 ANTIGENS OF MORAXELLA CATARRHALIS
FILE REFERENCE: AMCY.024
CURRENT APPLICATION NUMBER: US/09/952,267
PRIOR FILING DATE: 2001-09-12
PRIOR APPLICATION NUMBER: 09/336,447
PRIOR FILING DATE: 1999-06-21
NUMBER OF SEQ ID NOS: 98
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 13
LENGTH: 873
TYPE: PR1
ORGANISM: Moraxella catarrhalis
US-09-952-267-13

Query Match
Best Local Similarity 10.7%; Score 83.5; DB 10; Length 873;
Matches 40; Conservative 20; Mismatches 44; Indels 57; Gaps 10;

Cy 4 LKVAAPFAIVVSGSALAGVPPQWGGGNNH--GGGNS-----SGPDYDOLVTVRV 51
Db 34 LILVIGLIGMATTASAGQOTIARQ--GKGMSIIGGNDNEANGDYSTVSGGDYNE----- 85

Cy 52 THEMAHALGSDARKSETTITQSGYNGADVGADGADNSTBELTONGFRNNATIDOMAKNS 111
Db 86 -----AKGDSSTIGGGYNEAN-----GDSSTI-----GCGFYN-----BAKGR 119

Cy 112 DITVGOYGGNNALVNTASDSSVMTRQVFG--NNATANGY 151
Db 120 SSTIG--GGDN-----NSATGMYSTIGGGDNNSATGRY 150

RESULT 14
US-10-238-075-1549
Sequence 1549, Application US/10238075
Publication No. US20030148324A1
GENERAL INFORMATION:
APPLICANT: I.N.S.E.R.M.
TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isolat
FILE REFERENCE: BLANDINE
CURRENT APPLICATION NUMBER: US/10/238,075

```

; CURRENT FILING DATE: 2002-09-10
; PRIOR APPLICATION NUMBER: 0003145
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 1576
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1549
; LENGTH: 182
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-238-075-1549

Query Match 10.6%; Score 83; DB 14; Length 182;
Best Local Similarity 24.2%; Pred. No. 1.2;
Matches 38; Conservative 26; Mismatches 77; Indels 16; Gaps 5;

QY 1 MKLKVAFAIIVSGSALA--GVVPQWGGGNNHGGSSGPDYDQVTRVYTHMAHA 58
DB 3 IKVIALATFVSAFVAGSMAVDGTTFTGKVAQCTVNTSDKDLAVTLPTVAT---SS 58
QY 59 LQSDARKSETT---ITSGYNGADVGGA----DNSTIELTQGFNNATIDQWNAKN 110
DB 59 LKDNATSGLTFFAIRLGCATGMSAQNVKAYFEPSNIDLATHNLKNTAT--PTKADN 116
QY 111 SDITVGQYGNNAALVNOTASDSSVMVRQVGFNNAT 147
DB 117 VQIQILNSNGTSTILLGEADGQDVOSETIGSDSAT 153

RESULT 15

US-09-820-843A-21
; Sequence 21, Application US/09820843A
; Publication No. US2003003963A1
; GENERAL INFORMATION:
; APPLICANT: Council of Scientific and Industrial Research
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEI
; FILE REFERENCE: Q63915
; CURRENT APPLICATION NUMBER: US/09/820,843A
; CURRENT FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 21
; LENGTH: 354
; TYPE: PRT
; ORGANISM: M. tuberculosis
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: PPE
; NAME/KEY: misc feature
; OTHER INFORMATION: 91|1781260
US-09-820-843A-21

Query Match 10.6%; Score 83; DB 10; Length 354;
Best Local Similarity 25.2%; Pred. No. 2.7;
Matches 31; Conservative 11; Mismatches 47; Indels 34; Gaps 3;

QY 27 GGGGNHNGGNSGPDYDQVTRVYTHMAHALQSDARKSETTITQSGYNGADVGAD 86
DB 46 GNNGNFNFSGSNTG-----SNNIGFNTGSGNFGFGNTGNN 81
QY 87 NSTIELTQGFNNATIDQWNAKSDITVGQYGNNAALVNOTASDSSVMVRQVGFNNA 146
DB 82 NIGIGLTGDC--QIGIGLNSGSGNIGFNSGTGNGLFNSGTGN-----VGFQNSG 131
QY 147 TAN 149
DB 132 TAN 134

Search completed: March 11, 2004, 19:18:37
Job time : 25.6 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:24:14 ; Search time 10.3 Seconds

(without alignments)
1410.186 Million cell updates/sec

Title: US-09-543-407-18

Perfect score: 780

Sequence: 1 MKLKVAAFPAAIVSGSALA.....DSSVMVROVGFGNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	691	88.6	151	2 JG6039	fimbria protein ag
2	691	88.6	151	2 A10635	major curlin chain
3	523	67.1	151	2 S70788	curlin protein csg
4	501.5	64.3	152	2 D90806	curlin major subun
5	501.5	64.3	152	2 H85665	cnjB protein - Tet
6	114	14.6	1748	2 S42136	curlin nucleator p
7	107.5	13.8	151	2 S70787	minor curlin subun
8	107.5	13.8	151	2 C90806	curlin minor chain
9	107.5	13.8	151	2 G85665	fimbria protein ag
10	106.5	13.7	151	2 UC6040	nucleation compone
11	106.5	13.7	151	2 A10635	conserved hypochet
12	95.5	12.2	145	2 AD3143	hypothetical prote
13	95.5	12.2	145	2 H98144	probable PPE prote
14	93.5	12.0	582	2 F70675	hypothetical prote
15	93	11.9	141	2 AB3143	hypothetical prote
16	93	11.9	141	2 B98145	hypothetical prote
17	92	11.8	401	2 C88571	protein C05B5.3 (1
18	91.5	11.7	590	1 A45621	leishmanolysin (EC
19	91.5	11.7	2174	2 E95965	hypothetical glyci
20	91	11.5	590	2 E70946	probable PPE prote
21	90	11.5	602	1 PL0221	leishmanolysin (EC
22	90	11.5	896	2 A41273	glutamate receptor
23	90	11.5	906	2 A40222	glutamate receptor
24	89.5	11.5	906	2 S25852	leishmanolysin (EC
25	89.5	11.5	599	2 B42049	hypothetical prote
26	88.5	11.3	1655	2 E97835	leishmanolysin (EC
27	88	11.3	599	2 A44951	glutamate receptor
28	88	11.3	906	2 S38723	glutamate receptor
29	87.5	11.2	552	2 D70604	probable PPE prote

30	87.5	11.2	1028	2 A56038	DNA-binding protei
31	87.5	11.2	1213	2 S16356	ovo protein - fru
32	87.5	11.2	1567	2 S1672	ice nucleation pro
33	87	11.2	407	2 T21956	hypothetical prote
34	86	11.0	409	2 T20847	hypothetical prote
35	86	11.0	495	2 B71360	hypothetical prote
36	85.5	11.0	588	2 F70971	hypothetical glyci
37	85	10.9	364	2 S43574	C05B5.3 protein (c
38	85	10.9	645	2 F70825	probable PPE prote
39	85	10.9	3716	2 E70969	probable PPE prote
40	85	10.9	5188	2 B85547	probable RTX fami
41	85	10.9	5291	2 F90966	hypothetical prote
42	84.5	10.8	639	2 C42049	leishmanolysin (EC
43	84	10.8	586	2 T26667	hypothetical prote
44	84	10.8	633	2 A25473	chorion E2 protein
45	84	10.8	646	1 S19916	leishmanolysin (EC

ALIGNMENTS

RESULT 1

JG6039 fimbria protein agfa precursor - Salmonella enteritidis

C/Species: Salmonella enteritidis

C/Date: 31-Dec-1996 #sequence revision 31-Dec-1996 #text_change 08-Oct-1999

C/Accession: JG6039; PC6015; A44898

R/Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Baner, P.A.; Kay, W.W.

J. Bacteriol. 178, 662-667, 1996

A/Title: Salmonella enteritidis agfaAC operon encoding thin, aggregative fimbriae.

A/Reference number: JG6039; MUID:96146512; PMID:8550497

A/Accession: JG6039

A/Molecule type: DNA

A/Residues: 1-151 <COL>

A/Cross-references: GB:U43280; NID:G1184712; PIDN:AAC3599.1; PID:G1184714

A/Accession: PC6015

A/Molecule type: protein

A/Residues: 21-52 <CO2>

A/Experimental source: strain 27655-3b

A/Note: the authors translated the codon ACG for residue 44 as Ile

R/Collinson, S.K.; Emdy, L.; Muller, K.H.; Trust, T.J.; Kay, W.W.

J. Bacteriol. 173, 4773-4781, 1991

A/Title: Purification and characterization of thin, aggregative fimbriae from Salmonella

A/Reference number: A44898; MUID:91310586; PMID:1677357

A/Accession: A44898

A/Contents: 27655

A/Status: preliminary

A/Molecule type: protein

A/Residues: 21-33 <CO3>

A/Note: sequence extracted from NCBI backbone (NCBIP:45936)

C/Genetics:

A/Gene: agfa

C/Function:

A/Description: major component of thin aggregative fimbriae

A/Note: fimbriae bind to fibronectin, plasminogen, tissue plasminogen activator

C/Keywords: fimbria

F/1-20/Domain: signal sequence #status predicted <SIG>

F/21-151/Product: fimbria protein agfa #status experimental <MAT>

Query Match

Best Local Similarity 90.7%; Pred. No. 9.7e-53;

Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY	1	MKLKVAAFPAAIVSGSALAGVPPMGGGGNNNGGNSGPPYDQLVTRVTHEMAHALQ	60
DB	1	MKLKVAAFPAAIVSGSALAGVPPMGGGGNNNGGNSGPPSTLSIYYGSANALALQ	60
QY	61	SDARKSETTITGSGNGADVQAGADNSTIELTONGFRNNATIDQNNANSDITVQYGG	120
DB	61	SDARKSETTITGSGNGADVQAGADNSTIELTONGFRNNATIDQNNANSDITVQYGG	120
QY	121	NNAAVNGTASDSSVMVROVGFGNATANQY 151	
DB	121	NNAAVNGTASDSSVMVROVGFGNATANQY 151	

Db 121 NNAALVNQTASDSSVMVROVGFNNATANQY 151

RESULT 2

major curlin chain precursor [imported] - Salmonella enterica subsp. enterica serovar Typhimurium
 C/Species: Salmonella enterica subsp. enterica serovar Typhimurium
 A/Note: this species has also been called Salmonella typhimurium
 C/Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
 C/Accession: A10635
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Churcher, M.; T. Comerston, P.; Cronin, A.; Davis, P.; Davies, K.M.; Dowd, L.; White, N.; Farrar, S.; Mout, S.; O'Garra, P.
 Nature 413, 848-852, 2001
 A/Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A/Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhimurium
 A/Reference number: AB0502; MUID:21534947; PMID:11677608
 A/Accession: A10635
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-151 <PAR>
 A/Cross-references: GB:AL513382; PIDN:CAD08268.1; PID:g16502315; GSPDB:GN00176
 C/Genetics:
 A/Gene: STY1181

Query Match 88.6%; Score 691; DB 2; Length 151;
 Best Local Similarity 90.7%; Pred. No. 9.7e-53;
 Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 1 MLLKVAFAAIVVSSALAGVVPQWGGGNNHGGSSGPDYDQVTVRVVTHMAHALQ 60
 Db 1 MLLKVAFAAIVVSSALAGVVPQWGGGNNHGGSSGPDYDQVTVRVVTHMAHALQ 60
 QY 61 SDARKSETTTTOSGYGNGADVGGADNSTIELTONGFRNNATIDOMNAKNSDITVGGYGG 120
 Db 61 SDARKSETTTTOSGYGNGADVGGADNSTIELTONGFRNNATIDOMNAKNSDITVGGYGG 120
 QY 121 NNAALVNQTASDSSVMVROVGFNNATANQY 151
 Db 121 NNAALVNQTASDSSVMVROVGFNNATANQY 151

RESULT 3

curlin protein csgA precursor - Escherichia coli (strain K-12)
 N/Alternate names: csgA protein; major curlin protein
 C/Species: Escherichia coli
 C/Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 01-Mar-2002
 C/Accession: S70788; G64846; S31202; S34560; S34559
 R:Hammar, M.; Arngvist, A.; Bian, Z.; Olsen, A.; Normark, S.
 Mol. Microbiol. 18, 661-670, 1995
 A/Title: Expression of two csg operons is required for production of fibronectin- and Cc
 A/Reference number: S70783; MUID:96414468; PMID:8817489
 A/Accession: S70788
 A/Status: nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-151 <HAM>
 A/Cross-references: EMBL:X90754; NID:g1147558; PIDN:CAA62282.1; PID:g1147564
 A/Experimental source: strain K12, substrain W3110
 A/Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
 A.; Rose, D.; Mau, B.; Sha, Y.
 Science 277, 1453-1462, 1997
 A/Title: The complete genome sequence of Escherichia coli K-12.
 A/Reference number: A64720; MUID:97426617; PMID:9278503
 A/Accession: G64846
 A/Status: nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-151 <BIAT>
 A/Cross-references: GB:AB00205; GB:U00096; NID:g1787265; PIDN:AACT4126.1; PID:g1787279;
 A/Experimental source: strain K-12, substrain MG1655
 R:Olson, A.; Arngvist, A.; Hammar, M.; Sukupolvi, S.; Normark, S.
 Mol. Microbiol. 7, 523-536, 1993

A/Title: The RpoS sigma factor relieves H-NS-mediated transcriptional repression of csgA
 A/Reference number: S31202; MUID:93211294; PMID:8459772
 A/Accession: S31202

A/Molecule type: DNA
 A/Residues: 1-6, 'V', 8-151 <OL51>
 A/Cross-references: EMBL:U04979
 A/Accession: S34560
 A/Molecule type: protein
 A/Residues: 21-42;44-50 <OL52>
 R:Olson, A.N.; Arngvist, A.M.
 submitted to the EMBL Data Library, October 1992
 A/Reference number: S34559
 A/Accession: S34559
 A/Molecule type: DNA
 A/Residues: 1-133, 'RORSGLW' <OL53>
 A/Cross-references: EMBL:U04979; NID:g290424; PIDN:AAA23616.1; PID:g290425
 A/Experimental source: strain K-12, substrain W3110
 C/Genetics:
 A/Gene: csgA
 A/Map position: 23.15
 C/Function:
 A/Description: major component of wild-type curli; interaction between CsgA and CsgB tr
 A/Note: curli are thin, coiled fibers expressed on the surface of Escherichia coli that
 and H-kinogen; in the absence of CsgA, CsgB can self-assemble into polymers
 F;1-20/Domain: signal sequence #status predicted <SIS>
 F;21-151/Product: curlin #status experimental <MAT>

Query Match 67.1%; Score 523; DB 2; Length 151;
 Best Local Similarity 68.9%; Pred. No. 3.1e-38;
 Matches 104; Conservative 20; Mismatches 27; Indels 0; Gaps 0;

QY 1 MLLKVAFAAIVVSSALAGVVPQWGGGNNHGGSSGPDYDQVTVRVVTHMAHALQ 60
 Db 1 MLLKVAFAAIVVSSALAGVVPQWGGGNNHGGSSGPDYDQVTVRVVTHMAHALQ 60
 QY 61 SDARKSETTTTOSGYGNGADVGGADNSTIELTONGFRNNATIDOMNAKNSDITVGGYGG 120
 Db 61 SDARKSETTTTOSGYGNGADVGGADNSTIELTONGFRNNATIDOMNAKNSDITVGGYGG 120
 QY 121 NNAALVNQTASDSSVMVROVGFNNATANQY 151
 Db 121 NNAALVNQTASDSSVMVROVGFNNATANQY 151

RESULT 4

curlin major subunit CsgA [imported] - Escherichia coli (strain O157:H7, substrain RIMD
 C/Species: Escherichia coli
 C/Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
 C/Accession: D90806
 R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G
 gasawara, N.; Yasunaga, T.; Kohara, S.; Shiba, T.; Hattori, M.; Shingawa, H.
 DNA Res. 8, 11-22, 2001
 A/Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen
 A/Reference number: A99629; MUID:21156231; PMID:11258796
 A/Accession: D90806
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-152 <HAY>
 A/Cross-references: GB:BA00007; PIDN:BAB34843.1; PID:g13360880; GSPDB:GN00154
 A/Experimental source: strain O157:H7, substrain RIMD 0509952
 C/Genetics:
 A/Gene: ECE1420

Query Match 64.3%; Score 501.5; DB 2; Length 152;
 Best Local Similarity 67.1%; Pred. No. 2.3e-36;
 Matches 102; Conservative 21; Mismatches 28; Indels 1; Gaps 1;

QY 1 MLLKVAFAAIVVSSALAGVVPQWGGGNNHGGSSGPDYDQVTVRVVTHMAHALQ 59
 Db 1 MLLKVAFAAIVVSSALAGVVPQWGGGNNHGGSSGPDYDQVTVRVVTHMAHALQ 60
 QY 60 QSDARKSETTTTOSGYGNGADVGGADNSTIELTONGFRNNATIDOMNAKNSDITVGGYGG 119

Db 61 QADARSDLTITQHGGGNGADVGQSSDDSIDLTQGFNGSATLDDQNGKSHMTVKQFG 120

Qy 120 GNNALVNOTASDSSVMVRQVFGNNATANQY 151

Db 121 GGGAAVDDQTASNTSTVNTVQVFGNNATANQY 152

RESULT 5

H85665

hypothetical protein csgA [imported] - Escherichia coli (strain O157:H7, substrain EDL93

C/Species: Escherichia coli

C/Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #ext_change 14-Sep-2001

C/Accession: H85665

R/Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.U.; Mayhew

iller, L.; Grobbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,

Nature 409, 529-533, 2001

A/Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A/Reference number: AB5480; MUID:21074935; PMID:11206551

A/Accession: H85665

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-152 <STO>

A/Cross-references: GB:AE005174; NID:q12514574; PIDN:AA655788.1; GSPDB:GN00145; UMGF:216

A/Experimental source: strain O157:H7, substrain EDL933

C/Genetics:

A/Gene: csgA

Query Match 64.3%; Score 501.5; DB 2; Length 152;

Best Local Similarity 67.1%; Pred. No. 2.3e-36;

Matches 102; Conservative 21; Mismatches 28; Indels 1; Gaps 1;

Qy 1 MKLLKVAAPAAIVVSGSALAGVPPQW-GGGGNNHGGNSGPDYDQVTRVVTTHMAHAL 59

Db 1 MKLLKVAALIAIVFSSGALAGVPPQYGGGGNNHGGNSGPPSELNITVQYGGGALALL 60

Qy 60 QSDARSETTTTQSGYGGADVGQGDNSTIELTQNGFRNNATTIDQMAKNSDITVQYQ 119

Db 61 QADARSDLTITQHGGGNGADVGQSSDDSIDLTQGFNGSATLDDQNGKSHMTVKQFG 120

Qy 120 GNNALVNOTASDSSVMVRQVFGNNATANQY 151

Db 121 GGGAAVDDQTASNTSTVNTVQVFGNNATANQY 152

RESULT 6

S42136

cnbB protein - Tetrahymena thermophila

C/Species: Tetrahymena thermophila

C/Date: 19-Mar-1997 #sequence_revision 18-Jul-1997 #ext_change 07-Dec-1999

C/Accession: S42136; S42135; S03650

R/Taylor, F.M.; Martindale, D.W.

submitted to the EMBL Data Library, October 1992

A/Reference number: S42136

A/Accession: S42136

A/Molecule type: DNA

A/Residues: 1-1748 <TAY>

A/Cross-references: EMBL:U03710; NID:g161751; PID:g161752

R/Taylor, F.M.; Martindale, D.W.

Nucleic Acids Res. 21, 4610-4614, 1993

A/Title: Retroviral-type zinc fingers and glycine-rich repeats in a protein encoded by c

A/Reference number: S42135; MUID:94051569; PMID:8233798

A/Accession: S42135

A/Molecule type: DNA

A/Residues: 1164-1174; 1179-1198; 1233-1252; 1285-1293; 1297-1309; 1316-1326; 1331-1341; 1343-1

A/Cross-references: EMBL:U03710

R/Martindale, D.W.; Taylor, F.M.

Nucleic Acids Res. 16, 2189-2201, 1988

A/Title: Multiple functions in a conjugation-specific gene from Tetrahymena thermophila.

A/Reference number: S03650; MUID:88189811; PMID:3357771

A/Accession: S03650

A/Molecule type: DNA

A/Residues: 236-250; 'T', 252-255; 'N', 257-773 <MAR>

A/Cross-references: EMBL:X06462

C/Genetics:

A/Gene: cnbB

A/Genetic code: SGC5

A/Introns: 85/3; 136/1; 157/3; 201/2; 290/2; 327/3; 499/1; 573/2; 607/3; 708/3; 777/3;

C/Keywords: zinc finger

F/1164-1450/Region: glycine-rich

F/1451-1464/Region: zinc finger CCHC motif

F/1478-1491/Region: zinc finger CCHC motif

F/1501-1514/Region: zinc finger CCHC motif

F/1530-1543/Region: zinc finger CCHC motif

F/1555-1568/Region: zinc finger CCHC motif

F/1579-1592/Region: zinc finger CCHC motif

F/1602-1615/Region: zinc finger CCHC motif

F/1626-1748/Region: glycine-rich

Query Match 14.6%; Score 114; DB 2; Length 1748;

Best Local Similarity 30.5%; Pred. No. 0.11; 27; Indels 40; Gaps 8;

Matches 39; Conservative 22; Mismatches 27; Indels 40; Gaps 8;

Qy 25 OMGGGNNHGG---GNSGPDYDQVTRVVTTHMAHALQSDARSETTTTQSGYGGADV 81

Db 1640 QFGGGGNSNGGSGWGTSSGSDMN-----CGSNVQES-TTSSSGWGS----- 1680

Qy 82 QGQADNSTIELTQNGFRNNATTIDQMAKNSDITVQYGGNNAALVNOTASDSSVMVRQV 141

Db 1681 -SGSGNQ---TGGGWSN---DNOQQNENTGGGWSNS---NOTNNES----- 1722

Qy 142 FGNNATAN 149

Db 1723 WGSNNQAS 1730

RESULT 7

S70787

curlin nucleator protein csgB precursor - Escherichia coli (strain K-12)

N/Alternate names: csgB protein; curlin nucleation component; minor curlin protein

C/Species: Escherichia coli

C/Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #ext_change 01-Mar-2002

C/Accession: S70787; F64846

R/Hammar, M.; Arngvist, A.; Bian, Z.; Olsen, A.; Normark, S.

Mol. Microbiol. 18, 661-670, 1995

A/Title: Expression of two csg operons is required for production of fibronectin- and C

A/Reference number: S70783; MUID:96414468; PMID:8817489

A/Accession: S70787

A/Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-151 <HAM>

A/Cross-references: EMBL:X90754; NID:g1147558; PIDN:CA62281.1; PID:g1147563

A/Experimental source: strain K12, substrain W3110

A/Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995

R/Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C

A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A/Title: The complete genome sequence of Escherichia coli K-12.

A/Reference number: A64720; MUID:97426617; PMID:9278503

A/Accession: F64846

A/Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-151 <BLAT>

A/Cross-references: GB:AE000205; GB:U00096; NID:g1787265; PIDN:AACT4125.1; PID:g1787278,

A/Experimental source: strain K-12, substrain MG1655

C/Genetics:

A/Gene: csgB

A/Map position: 23.15

C/Function:

A/Description: minor component of wild-type curli; interaction between CsgA and CsgB tr

A/Note: curli are thin, coiled fibers expressed on the surface of Escherichia coli that

and H-kinogenin in the absence of CsgA, CsgB can self-assemble into polymers

F/1-21/Domain: signal sequence #status predicted <SIG>

F/22-151/Product: minor curlin chain #status predicted <MAT>

Query Match 13.8%; Score 107.5; DB 2; Length 151;

Best Local Similarity 35.4%; Pred. No. 0.024;
Matches 29; Conservative 9; Mismatches 43; Indels 1; Gaps 1;
QY 70 ITGSGYGNADVCGADNSTIELTONGFRNNATTIDQNNAKNSDITVGYGNNALVNOT 129
DB 49 IGGAGTNNSAQLRGGSKLAVVAQEGSSNRAKIDGTGYVL-AYIDQGSANDASISOG 107
QY 130 ASDSSVMVROVGFNNATANQY 151
DB 108 AYGNMTAMIIQKSGSKANITQY 129

RESULT 8
C90806
minor curlin subunit precursor CsgB [imported] - Escherichia coli (strain O157:H7, subsp
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: C90806
R:Hayashi, T.; Makino, K.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene
A:Reference number: A59629; MUID:21156231; PMID:11258796
A:Accession: C90806
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-151 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA834842.1; PID:G13360879; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain R1MD 050952
C:Genetics:
A:Gene: EC81419

Query Match 13.8%; Score 107.5; DB 2; Length 151;
Best Local Similarity 35.4%; Pred. No. 0.024;
Matches 29; Conservative 9; Mismatches 43; Indels 1; Gaps 1;
QY 70 ITGSGYGNADVCGADNSTIELTONGFRNNATTIDQNNAKNSDITVGYGNNALVNOT 129
DB 49 IGGAGTNNSAQLRGGSKLAVVAQEGSSNRAKIDGTGYVL-AYIDQGSANDASISOG 107
QY 130 ASDSSVMVROVGFNNATANQY 151
DB 108 AYGNMTAMIIQKSGSKANITQY 129

RESULT 9
G85665
curlin minor chain precursor, CsgA homolog [imported] - Escherichia coli (strain O157:H7
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: G85665
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: G85665
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-151 <STO>
A:Cross-references: GB:AB005174; NID:G12514573; PIDN:AA655787.1; GSPDB:GN00145; UMGF:216
C:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: csgB

Query Match 13.8%; Score 107.5; DB 2; Length 151;
Best Local Similarity 35.4%; Pred. No. 0.024;
Matches 29; Conservative 9; Mismatches 43; Indels 1; Gaps 1;
QY 70 ITGSGYGNADVCGADNSTIELTONGFRNNATTIDQNNAKNSDITVGYGNNALVNOT 129
DB 49 IGGAGTNNSAQLRGGSKLAVVAQEGSSNRAKIDGTGYVL-AYIDQGSANDASISOG 107

QY 130 ASDSSVMVROVGFNNATANQY 151
DB 108 AYGNMTAMIIQKSGSKANITQY 129

RESULT 10
JC6040
fimbria protein agfB precursor - Salmonella enteritidis
C:Species: Salmonella enteritidis
C:Date: 31-Dec-1996 #sequence_revision 31-Dec-1996 #text_change 08-Oct-1999
C:Accession: JC6040
R:Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Baner, P.A.; Kay, W.W.
J. Bacteriol. 178, 662-667, 1996
A:Title: Salmonella enteritidis agfBAC operon encoding thin, aggregative fimbriae.
A:Reference number: JC6039; MUID:96146512; PMID:8550497
A:Accession: JC6040
A:Molecule type: DNA
A:Residues: 1-151 <COL>
A:Cross-references: GB:U43280; NID:G1184712; PIDN:AA043598.1; PID:G1184713
A:Experimental source: strain 276755-3b
C:Genetics:
A:Gene: agfB
C:Function:
A:Description: minor component of thin aggregative fimbriae
A>Note: fimbriae bind to fibronectin, plasminogen, tissue plasminogen activator
C:Keywords: fimbria
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-151/Product: fimbria protein agfB #status predicted <MAT>

Query Match 13.7%; Score 106.5; DB 2; Length 151;
Best Local Similarity 30.7%; Pred. No. 0.029;
Matches 35; Conservative 17; Mismatches 43; Indels 19; Gaps 5;
QY 55 MAHALQSDARKSE-----TTTQSGYGNADVCGADNST-----IELTONGFR 98
DB 18 IATATNYDLARSEYNFAVNELSKSSFNQAAIIGVGTNDSARVROGSKLSVISOEGSN 77
QY 99 NNATTIDQNNAKNSDIT-VGYGNNALVNOTASVVMVROVGFNNATANQY 151
DB 78 NNAKYDQ--AGNYNFAVIEQTGNANDASISQASNGNSAAIIQKSGSKANITQY 129

RESULT 11
AH0635
nucleation component of curlin monomers [imported] - Salmonella enterica subsp. enteric
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A>Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AH0635
R:Parhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher
th, T.; Conerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica sero
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AH0635
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-151 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD08267.1; PID:G16502314; GSPDB:GN00176
C:Genetics:
A:Gene: STY1180

Query Match 13.7%; Score 106.5; DB 2; Length 151;
Best Local Similarity 30.7%; Pred. No. 0.029;
Matches 35; Conservative 17; Mismatches 43; Indels 19; Gaps 5;
QY 55 MAHALQSDARKSE-----TTTQSGYGNADVCGADNST-----IELTONGFR 98
DB 18 IATATNYDLARSEYNFAVNELSKSSFNQAAIIGVGTNDSARVROGSKLSVISOEGSN 77
QY 99 NNATTIDQNNAKNSDIT-VGYGNNALVNOTASVVMVROVGFNNATANQY 151

Db 78 NRAKVDQ--AGNVPAYIEQTGNANDASISQSAVGNSSAIIQSGSNKANTQY 129

RESULT 12

AD3143

Conserved hypothetical protein Atu4768 [imported] - Agrobacterium tumefaciens (strain C58)
C/Species: Agrobacterium tumefaciens
C/Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C/Accession: AD3143

R.Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.
erage, G.; Gallet, W.; Grant, C.; Guenther, D.; Kutyavyn, T.; Levy, R.; Li, M.; McClellan,
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001

A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A/Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A/Reference number: AB2577; MUID:21608550; PMID:11743193

A/Accession: AD3143

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-145 <RUR>

A/Cross-references: GB:AE008689; PIDN:AA145562.1; PID:g17743277; GSPDB:GN00187

A/Experimental source: strain C58 (Dupont)

C/Genetics:

A/Map position: linear chromosome

A/Gene: Atu4768

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

QY 50 VTHEMAHALQSPARKSETTITGSGYNGADVGQADNSTITLTONGFNNATTIDQMAK 109
Db 61 IVGH-----QYGRHNLSAVGQEGHDVYGSTTQGNRNVAIGI----- 96

RESULT 14

F70675

probable PPE protein - Mycobacterium tuberculosis (strain H37Rv)
C/Species: Mycobacterium tuberculosis
C/Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 15-Sep-2003
C/Accession: F70675

R.Cole, S.T.; Brosch, R.; Parthill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon,
; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamilton, N.; Holtroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998

A/Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A/Reference number: A70500; MUID:98295987; PMID:9634230

A/Accession: F70675
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA

A/Residues: 1-582 <COL>
A/Cross-references: GB:Z82098; GB:AL123456; NID:9326164; PIDN:CAB05045.1; PID:e280737;

A/Experimental source: strain H37Rv
C/Genetics:

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

QY 110 NSDITVGOYGNNALVNQTAADSSVMVROVGFNNATTANQ 150
Db 97 -----GQFGSNHTTITLTDGNGNIAAGVQVGRGCSANVSQ 131

RESULT 15

AB3143

hypothetical protein Atu4766 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C/Species: Agrobacterium tumefaciens
C/Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C/Accession: AB3143

R.Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.
erage, G.; Gallet, W.; Grant, C.; Guenther, D.; Kutyavyn, T.; Levy, R.; Li, M.; McClellan,
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001

A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A/Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A/Reference number: AB2577; MUID:21608550; PMID:11743193

A/Accession: AB3143
A/Status: preliminary
A/Molecule type: DNA

A/Residues: 1-141 <RUR>
A/Cross-references: GB:AE008689; PIDN:AA145560.1; PID:g17743275; GSPDB:GN00187

A/Experimental source: strain C58 (Dupont)
C/Genetics:

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

A/Map position: linear chromosome

Best Local Similarity 26.0%; Pred. No. 0.39;
Matches 39; Conservative 22; Mismatches 63; Indels 26; Gaps 6;

```

QY      1 MKLIKVAFAIIVSGSALAGVVPONGGGGNHNG-----GNSSGPDYDQLVTRVVT 52
      :|||
      :|||
Db      10 IKLITVTIL-----VGGLTPLGFAAPAHAAGGISFDIAPGNNA--DGDLLSTGL-- 56
      :|||
QY      53 HEMAHALOSDARKSETTITOSGYNGADVGQGDADNSTIELTONGFRNNATIDQWNAKNSD 112
      :|||
Db      57 --RAYSLYRNPK--DANIKQUGRGNMAAGIAONGGNIIGFTIRGRNGHSATLQQ--NGNNNA 111
      :|||
QY      113 ITVGQYGGNNAALVNOTASDSSVVRQVGF 142
      :|||
Db      112 YGIFQYGRNTGTINVVQDDGNGSGLTFSYGW 141
      :|||

```

Search completed: March 11, 2004, 18:42:09
Job time : 11.3 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Comugen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:14:48 ; Search time 6.3 Seconds
(without alignments)
1248.031 Million cell updates/sec

Title: US-09-543-407-18

Perfect score: 780
Sequence: 1 MLLKLVAFAPAIIVSGSALA.....DSSVMRYQFGNNATANY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	691	88.6	151	1 CSGB_SALTY	P55225 salmonella
2	523	67.1	151	1 CSGB_ECOLI	P28307 escherichia
3	501.5	64.3	152	1 CSGB_ECO57	Q93424 escherichia
4	107.5	13.8	151	1 CSGB_ECOLI	P39828 escherichia
5	106.5	13.7	151	1 CSGB_SALTY	O87263 salmonella
6	106.5	13.7	151	1 CSGB_SALTY	P55226 salmonella
7	92	11.8	401	1 YK03_CABEL	P34291 caenorhabdi
8	92	11.8	1656	1 OMPB_RICUA	O06653 r outer mem
9	91.5	11.7	590	1 GPE3_LEIDO	P23223 leishmania
10	91	11.7	1093	1 GPE3_LEIDO	O03397 leishmania
11	90	11.5	602	1 GPE3_LEIMA	P08148 leishmania
12	90	11.5	906	1 GLR1_HUMAN	P42261 homo sapien
13	88.5	11.3	1655	1 OMPB_RICAN	O9K43 r outer mem
14	88	11.3	589	1 GPE3_LEICH	P15706 leishmania
15	87.5	11.2	1028	1 OVO_DROME	P51521 drosophila
16	87.5	11.2	1567	1 ICEN_XNACT	P18127 xanthomonas
17	86	11.0	485	1 Y136_TREPA	O83172 treponema p
18	86	11.0	493	1 GATA_RHIME	O92067 rhizobium m
19	84	10.8	646	1 GPE3_LEIME	P43150 leishmania
20	84	10.8	907	1 GLR1_MOUSE	P23818 mus muscula
21	84	10.8	907	1 GLR1_MOUSE	P19194 rattus norv
22	83	10.6	720	1 KR6E_YEAST	P32486 saccharomyc
23	83	10.6	947	1 SECA_STRCO	P55001 streptomyce
24	82	10.5	342	1 OMPB_RAHQA	O33507 rahnella aq
25	81.5	10.4	424	1 COA4_BPPD	P03661 bacterioph
26	81.5	10.4	424	1 COA4_BPPD	P03662 bacterioph
27	81.5	10.4	493	1 GATP_AGRTS	Q8ute8 agrobacteri
28	81	10.4	365	1 ROAI_DROME	Q07909 drosophila
29	81	10.4	576	1 DEAF_DROME	Q24180 drosophila
30	81	10.4	678	1 YF48_MYCTU	Q10778 mycobacteri
31	81	10.4	1034	1 ICEN_PANAN	Q47893 panoea ana
32	80.5	10.3	1641	1 IMD_ARTGO	Q44052 arthrobaete
33	80.5	10.3	1185	1 MAPX_DROME	P23226 drosophila

ALIGNMENTS

RESULT 1	CSGB_SALTY	STANDARD	PRT	151 AA	003646 plasmodium
AC	P55225;				P95759 streptomyce
DT	01-OCT-1996 (Rel. 34, Created)				P14914 ticketsia
DT	01-OCT-1996 (Rel. 34, Last sequence update)				Q53047 r outer mem
DT	10-OCT-2003 (Rel. 42, Last annotation update)				P26968 tenebrio mo
DE	Major curilin subunit precursor (Fimbrin SEF17).				O04893 saccharomyc
GN	CSGB OR AGFA OR STM1144 OR STY1181 OR T1776.				P57037 neisseria m
OS	Salmonella typhimurium,				P24785 drosophila
OS	Salmonella typhi, and				P75139 mycoplasma
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;				P43096 candida alb
OC	Enterobacteriaceae; Salmonella.				P16239 erwina her
OX	NCHI_TaxID=602, 601, 592;				P33666 escherichia
RP	[1]				
RP	SEQUENCE FROM N.A.				
RC	SPECIES=S.typhimurium; STRAIN=SR-11;				
RA	MEDLINE=98117058; PubMed=9457880;				
RT	Romling U., Bian Z., Hammar M., Sierralta W.D., Normark S.;				
RT	"Curli fibers are highly conserved between Salmonella typhimurium and				
RT	Escherichia coli with respect to operon structure and regulation.";				
RL	J. Bacteriol. 180:722-731(1998).				
RP	[2]				
RP	SEQUENCE FROM N.A.				
RC	SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;				
RA	MEDLINE=21534948; PubMed=11677609;				
RA	McCllelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,				
RA	Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,				
RA	Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,				
RA	Ryan B., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,				
RA	Waterston R., Wilson R.K.;				
RT	"Complete genome sequence of Salmonella enterica serovar Typhimurium				
RT	LT2.";				
RL	Nature 413:852-856(2001).				
RP	[3]				
RP	SEQUENCE FROM N.A.				
RC	SPECIES=S.typhi; STRAIN=CT18;				
RA	MEDLINE=21534947; PubMed=11677608;				
RA	Parthill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,				
RA	Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,				
RA	Baker S., Basham D., Brooks K., Chillingworth T., Conerton P.,				
RA	Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrer J.,				
RA	Felwell T., Hamlin N., Haque A., Hien T.T., Holtroyd S., Jagels K.,				
RA	Krogg A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,				
RA	Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,				
RA	Whitehead S., Barrett B.G.;				
RT	"Complete genome sequence of a multiple drug resistant Salmonella				
RT	enterica serovar Typhi CT18.";				
RL	Nature 413:848-852(2001).				
RP	[4]				
RP	SEQUENCE FROM N.A.				
RC	SPECIES=S.typhi; STRAIN=Ty2 / ATCC 700931;				
RX	MEDLINE=22531367; PubMed=12644504;				
RX	Deng W., Lion S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,				
RA	Burford V., Kodoyanni V., Schwartz D.C., Blattner F.R.;				

"Comparative genomics of *Salmonella enterica* serovar Typhimurium strains Ty2 and CT18.";
 RL J. Bacteriol. 185:2330-2337(2003).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S. enteritidis; STRAIN=27655-3B;
 RX MEDLINE=96146512; PubMed=8550497;
 RA Collinson S.K., Cloutier S.C., Doran J.L., Baner P.A., Kay W.W.;
 RT "Salmonella enteritidis agfBAC operon encoding thin, aggregative
 fimbriae.";
 RL J. Bacteriol. 178:662-667(1996).
 RN [6]
 RP SEQUENCE OF 21-151 FROM N.A.
 RC SPECIES=S. enteritidis; STRAIN=27655-3B;
 RX MEDLINE=94013373; PubMed=8104955;
 RA Doran J.L., Collinson S.K., Burian J., Santos G., Todd E.C.D.,
 RA Munro C.K., Kay C.W., Baner P.A., Peterkin P.I., Kay W.W.;
 RT "DNA-based diagnostic tests for *Salmonella* species targeting agfA,
 the structural gene for thin, aggregative fimbriae.";
 RL J. Clin. Microbiol. 31:2263-2273(1993).
 RN [7]
 RP SEQUENCE OF 21-33.
 RC SPECIES=S. enteritidis; STRAIN=27655-3B;
 RX MEDLINE=91310586; PubMed=1677357;
 RA Collinson S.K., Emeody L., Mueller K.-M., Trust T.J., Kay W.W.;
 RT "Purification and characterization of thin, aggregative fimbriae from
Salmonella enteritidis.";
 RL J. Bacteriol. 173:4773-4781(1991).
 CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 CC FIBRONECTIN.
 CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements> /
 CC or send an email to license@ebi.ac.uk).
 CC -----
 DR EMBL; AJ002301; CAA05317.1; -
 DR EMBL; AE008749; AAL20074.1; -
 DR EMBL; AL627269; CAD08268.1; -
 DR EMBL; AE016840; AAC069399.1; -
 DR EMBL; U43280; AAC43599.1; -
 DR PIR; JC6039; JCC6039
 DR StyGene; SG10608; csgA.
 DR Fimbrin; Signal; Complete proteome.
 KW SIGNAL
 FT CHAIN 1 20 MAJOR CURLIN SUBUNIT.
 FT SIGNAL 21 151 SVMYROYFGNNATANY -> DSYTYVAS (IN
 FT CONFLICT 134 151 REF.6).
 SQ SEQUENCE 151 AA; 15305 MW; B7DACDD16B621359 CRC64;
 Query Match 88.6%; Score 691; DB 1; Length 151;
 Best Local Similarity 90.7%; Pred. No. 5.3e-53;
 Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

RESULT 2
 ID CSGA_ECOLI STANDARD; PRT; 151 AA.
 AC P28307;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Major curlin subunit precursor.
 GN CSGA OR B1042.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OC NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / W3110;
 RX MEDLINE=93211294; PubMed=8459772;
 RA Olsen A., Arngvist A.;
 RT "The Rpos sigma factor relieves H-NS-mediated transcriptional
 repression of csgA, the subunit gene of fibronectin-binding curli in
Escherichia coli.";
 RL Mol. Microbiol. 7:523-536(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MC4100;
 RX MEDLINE=96414468; PubMed=8817489;
 RA Hammar M., Arngvist A., Bian Z., Olsen A., Normark S.;
 RT "Expression of two csg operons is required for production of
 fibronectin- and congo red-binding curli polymers in *Escherichia coli*
 K-12.";
 RL Mol. Microbiol. 18:661-670(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Blythe M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of *Escherichia coli* K-12.";
 RL Science 277:1234-1238(1997).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97061202; PubMed=8905232;
 RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
 RA Ikemoto K., Inada T., Itoh T., Kajinaka M., Kanai K., Kashimoto K.,
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
 RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horuchi T.;
 RT "A 718-kb DNA sequence of the *Escherichia coli* K-12 genome
 corresponding to the 12.7-28.0 min region on the linkage map.";
 RL DNA Res. 3:117-155(1996).
 RN [5]
 RP SEQUENCE OF 21-40.
 RC STRAIN=K12 / YMBL;
 RX MEDLINE=93023873; PubMed=1157528;
 RA Arngvist A., Olsen A., Pfeiffer J., Russell D.G., Normark S.;
 RT "The cti protein activates cryptic genes for curli formation and
 fibronectin binding in *Escherichia coli* HB101.";
 RL Mol. Microbiol. 6:2443-2452(1992).
 RN [6]
 RP SEQUENCE OF 21-31.
 RC MEDLINE=91310586; PubMed=1677357;
 RA Collinson S.K., Emeody L., Trust T.J., Kay W.W.;
 RT "Purification and characterization of thin, aggregative fimbriae from
Salmonella enteritidis.";
 RL J. Bacteriol. 173:4773-4781(1991).
 CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 CC FIBRONECTIN.


```

CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L04979; AAA23616.1; -
DR EMBL; X80754; CAA62282.1; -
DR EMBL; AE000205; AAC74126.1; -
DR EMBL; D90741; BAA35832.1; -
DR EMBL; D90742; BAA35840.1; -
DR PIR; S70788; S70788.
DR Ecocore; EC11489; CSGA.
KW Fimbrtia; Signal; Complete proteome.
FT SIGNAL 1 20
FT CHAIN 21 151 MAJOR CURLIN SUBUNIT.
FT CONFLICT 7 7 A -> B (IN REF. 1).
SQ SEQUENCE 151 AA; 15049 MW; C003470D208D395F CRC64;

Query Match 67.1%; Score 523; DB 1; Length 151;
Best Local Similarity 68.9%; Pred. No. 1.7e-38;
Matches 104; Conservative 20; Mismatches 27; Indels 0; Gaps 0;

QY 1 MLLKVAAPAAIVSGSALAGVPPQMGGGGNNHGGSGPDYDQLVTRVTHEMAHALQ 60
DB 1 MLLKVAAPAAIVFSSSALAGVPPQYGGGGNHGGSGNNSEINITYGGGNSALALQ 60
QY 61 SDARKSETTITQSGYNGADVGGAGDNSTIELTONGFRNNATIDQWNAKNSDITVGOYG 120
DB 61 TDARNSDLITTHGGGNGADVGGSDSDSIDLTQRFGNSATILDQNGKNSERTVQFGG 120
QY 121 NNALVNOTASDSSVMYRQVGFNNATPNQY 151
DB 121 GNGAADVQTASNSVTNVTVQFGNNATPAHQY 151

RESULT 3
CSGA_ECO57 STANDARD; PRT; 152 AA.
ID CSGA_ECO57
AC Q93U24;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Major curlin subunit precursor.
CSGA OR Z1676 OR ECS1420.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / ATCC 43895;
RX MEDLINE=21218556; PubMed=11319125;
RA Uhlisch G.A., Keen J.B., Elder R.O.;
RT "Mutations in the csgD promoter associated with variations in curli
RT expression in certain strains of Escherichia coli O157:H7.";
RL Appl. Environ. Microbiol. 67:2367-2370(2001).
[2]
RN SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoudis K.,
RA Apodaca J., Amutharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).

```

```

RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Ikeda T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,
RA Kihara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001)
CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN.
CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF275733; AAK53212.1; -
DR EMBL; AE005315; AAG55788.1; -
DR EMBL; AP002554; BAB34843.1; -
DR PIR; D90806; D90806.
DR PIR; H85665; H85665.
KW Fimbrtia; Signal; Complete proteome.
FT SIGNAL 1 20
FT CHAIN 21 152 MAJOR CURLIN SUBUNIT.
SQ SEQUENCE 152 AA; 15099 MW; EE3D2D94DDE91243 CRC64;

Query Match 64.3%; Score 501.5; DB 1; Length 152;
Best Local Similarity 67.1%; Pred. No. 1.2e-36;
Matches 102; Conservative 21; Mismatches 28; Indels 1; Gaps 1;

QY 1 MLLKVAAPAAIVSGSALAGVPPQMGGGGNNHGGSGPDYDQLVTRVTHEMAHAL 59
DB 1 MLLKVAAPAAIVFSSSALAGVPPQYGGGGNHGGSGNNSEINITYGGGNSALAL 60
QY 60 OSDARKSETTITQSGYNGADVGGAGDNSTIELTONGFRNNATIDQWNAKNSDITVGOYG 119
DB 61 QADANSDLTTHGGGNGADVGGSDSDSIDLTQRFGNSATILDQWNGKSHMTVTKQFG 120
QY 120 GNGAADVQTASNSVTNVTVQFGNNATPAHQY 151
DB 121 GNGAADVQTASNSVTNVTVQFGNNATPAHQY 152

RESULT 4
CSGB_ECOLI STANDARD; PRT; 151 AA.
ID CSGB_ECOLI
AC P39828;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Minor curlin subunit precursor.
CSGB OR B1041 OR Z1675 OR ECS1419.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 83334;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MC4100;
RX MEDLINE=96414468; PubMed=8817489;
RA Hammar M., Arqvist A., Bian Z., Olsen A., Normark S.;
RT "Expression of two csg operons is required for production of
RT fibronectin- and Congo red-binding curli polymers in Escherichia coli

```

```

RT RT K-12." ;
RN RN Mol. Microbiol. 18:661-670 (1995).
RN RN [2]
RN RN SEQUENCE FROM N.A.
RP RP STRAIN=K12 / MG1655;
RX RX MEDLINE=97426617; PubMed=9278503;
RA RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT RT "the complete genome sequence of Escherichia coli K-12." ;
RL RL science 277:1453-1474 (1997).
RN RN [3]
RN RN SEQUENCE FROM N.A.
RP RP STRAIN=K12;
RX RX MEDLINE=97061202; PubMed=8905232;
RA RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kani K., Kashimoto K.,
RA Kimura S., Kitegawa M., Makino K., Masuda S., Miki T., Mizubuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Samei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horinouchi T.;
RT RT A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT RT corresponding to the 12.7-28.0 min region on the linkage map." ;
RL RL DNA Res. 3:137-155 (1996).
RN RN [4]
RN RN SEQUENCE FROM N.A.
RP RP STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX RX MEDLINE=21074935; PubMed=11206551;
RA RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans F.S., Gregor J., Kirkpatrick H.A.,
RA Postal E., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Dworkin R., Hinkle G., Peterson J.D., Beatty M., Li W., Lioy P.,
RA Apodaca J., Antantharam T.S., Lin J., Yan G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7." ;
RL RL Nature 409:529-533 (2001).
RN RN [5]
RN RN SEQUENCE FROM N.A.
RP RP STRAIN=O157:H7 / RIMD 0509952;
RX RX MEDLINE=21156231; PubMed=11258796;
RA RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohnishi E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasaki C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT RT O157:H7 and genomic comparison with a laboratory strain K-12." ;
RL RL DNA Res. 8:11-22 (2001).
RN RN [6]
RN RN SEQUENCE OF 1-21 FROM N.A.
RP RP STRAIN=K12;
RX RX MEDLINE=95157246; PubMed=7854117;
RA RA Arringvist A., Olsen A., Normark S.;
RT RT "Sigma S-dependent growth-phase induction of the csbA promoter in
RT RT Escherichia coli can be achieved in vivo by sigma 70 in the absence
RL RL of the nucleoid-associated protein H-NS." ;
RN RN Mol. Microbiol. 13:1021-1032 (1994)
RN RN [7]
RN RN FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
RN RN COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
RN RN TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
RN RN FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
RN RN CURLIN MONOMERS.
RN RN [8]
RN RN SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
RN RN [9]
RN RN This SWISS-PROT entry is copyright. It is produced through a collaboration
RN RN between the Swiss Institute of Bioinformatics and the EMBL outstation -
RN RN the European Bioinformatics Institute. There are no restrictions on its
RN RN use by non-profit institutions as long as its content is in no way
RN RN modified and this statement is not removed. Usage by and for commercial
RN RN entities requires a license agreement (see http://www.isb-sdb.ch/announce/
RN RN or send an email to license@isb-sdb.ch).
RN RN EMBL; X90754; CAA62281.1; -.

```

DR	EMBL; AE000205; AAC74125.1; -
DR	EMBL; D90741; BAA35831.1; -
DR	EMBL; AF005315; AAG55787.1; -
DR	EMBL; AF002554; BAB34042.1; --
DR	PIR; C90806; C90806.
DR	PIR; G85665; G85665.
DR	PIR; S70787; S70787.
DR	EcoGene; EGI2621; csgB.
KW	Fimbria; Signal; Complete proteome.
FT	SIGNAL
FT	CHAIN
FT	POTENTIAL.
FT	MIMOR CURLIN SUBUNIT.
SQ	SEQUENCE
	151 AA; 15882 MW; BI8D26B3640148 CRC64;
	22 151
	1 21
	13.8%; Score 107.5; DB 1; Length 151;
	Best Local Similarity 35.4%; Pred. No. 0.012;
	Matches 29; Conservative 9; Mismatches 43; Indels 1; Gaps 1
OY	70 ITDSGYGNGADVGGQGDNSTELTONGERNNATIDONNAKNSDTIVYGQGNNALVNQT 129
Dd	49 IGAGGTNNNAOALRCGGSKLLAVVAEGSSNRAKIDGTQDYNLAYIDOGASNDADISOG 107
OY	130 ASDSYVNRQVGFGNNTATANQY 151
Dd	108 AYENTAMITIQSGNKKANITQY 129

CC	CSGB_SALT1	STANDARD;	PRT;	151 AA.
CC	CSGB_SALT1			
AC	0827M3;			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Minor curlin subunit precursor.			
GN	CSGB OR STY1180 OR T1777.			
OC	<i>Salmonella typhi</i> .			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OC	Enterobacteriaceae; Salmonella.			
OX	NCBI_TaxId=601;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CT18;			
RC	MEDLINE=21534947; PubMed=11677608;			
RA	Patrickill U., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,			
RA	Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,			
RA	Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,			
RA	Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,			
RA	Felwell T., Hamlin N., Haque N., Hien T.T., Holroyd S., Jagels K.,			
RA	Kroch A., Laren T.S., Leather S., Moulé S., O'Gea P., Parry C.,			
RA	Quail M.A., Rutherford K., Simmonds M., Skellon T., Stevens K.,			
RA	Whithead S., Barrett B.G.;			
RT	"Complete genome sequence of a multiple drug resistant <i>Salmonella</i>			
RT	<i>enterica</i> serovar Typhi CT18.";			
RL	Nature 413:848-852 (2001).			
RL	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Ty2 / ATCC 700931;			
RC	MEDLINE=22531367; PubMed=12644504;			
RX	Deng W., Liu S.-R., Plunkett G. II, Mayhew G.F., Rose D.J.,			
RA	Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;			
RT	"Comparative genomics of <i>Salmonella enterica</i> serovar Typhi strains Ty2			
RT	and CT18.";			
RL	J. Bacteriol. 185:2330-2337 (2003).			
CC	-1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE			
CC	COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH			
CC	TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO			
CC	FIBRINOCTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF			
CC	CURLIN MONOMERS.			
CC	-1- SIMILARITY: BELONGS TO THE CSGB/CSGB FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			

	CC	use by non-profit institutions as long as its content is in no way modified and this statement is not removed.
	CC	modified and this statement is not removed. Usage by other entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to licenses@isb-sib.ch).
	CC	-----
	DR	EMBL; AL672269; CAD08267.1; -
	DR	EBL; AE016840; AAC069400.1; -
	KW	Fimbria; Signal; Complete proteome.
	FT	SIGNAL 1 21 POTENTIAL. CHAIN 22 151 MINOR CURLIN SUBUNIT.
	SO	SEQUENCE 151 AA; 16254 MW; 161C5432EE573495 CRC64;

	Oy	Query March 13 7%; Score 106.5; DB 1; Length 151; Best local similarity 30.7%; Pred. No. 0.015;
	Db	Matches 35; Conservative 17; Mismatches 43; Indels 19; Gaps 5;
	Oy	55 MAHALQSDPARSE-----TTTGGGCGAGAVGO-GADNST-----IELTNGFR 98 : : : : : : : : : : : : : : : : 18 IATNTNDVLASEVNFANVELSKSPFNAALIGVGDTNKSARVRQEGSKLSVSQGEN 77
	Db	99 NNATIDQMNAKNSDIT-VGYGNNALVNQTASDSVMVRQVFGNANTAYQY 151 : : : : : : : : : : : : : : : : 78 NRAKVQD--AGNYNPAYIEIQGNANDIASISGSAYGNSMAIIQKSGKNANTITY 129
	RESULT 6	
	CSCG_SALT	STANDARD; PRT; 151 AA.
	ID CSCG_SALT	P55226;
	AC	01-OCT-1996 (Rel. 34, Created)
	DT	01-OCT-1996 (Rel. 34, Last sequence update)
	DT	28-FEB-2003 (Rel. 41, Last annotation update)
	DE	Minor curlin subunit precursor (Pimdrin SEF17 minor subunit).
	Gn	CSCG OR AGFB OR STM1143.
	OS	Salmonella typhimurium, and Salmonella enteritidis.
	OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
	OC	Enterobacteriaceae; Salmonella.
	OX	NCBI_TaxID=602, 592;
	RN	[1]
	RP	SEQUENCE FROM N.A.
	RC	SPECIES=S.typhimurium; STRAIN=SR-11;
	RX	MEDLINE=96117058; PubMed=9457880;
	RA	McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P., Courtnay L., Portollik S., Ali U., Dante M., Du F., Hou S., Layman D., Leonard S., Nguyen C., Scott K., Holmes A., Grawal N., Mulvaney E., Ryan B., Sun H., Florea L., Miller W., Stoneking T., Nhan M., Waterston R., Wilson R.K.;
	RA	"Complete genome sequence of Salmonella enterica serovar Typhimurium LT2.";
	RL	Nature 413:852-856(2001). [3]
	RN	SEQUENCE FROM N.A.
	RP	SPECIES=S.enteritidis; STRAIN=27655-JB;
	RX	MEDLINE=96146512; PubMed=8550497;
	RA	Collinson S.K., Clouthier S.C., Doran J.L., Baner P.A., Kay W.M.;
	RA	"Salmonella enteritidis agtBA operon encoding thin, aggregative fimbriae".
	RT	J. bacteriol. 178:662-667(1996).
	RL	-1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF CURLIN MONOMERS.
	CC	-1- SIMILARITY: BELONGS TO THE CSGA/CSCG FAMILY.

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; AJ002301; CAA05316.1; -
CC DR EMBL; AE008749; AAL20073.1; -
CC DR EMBL; U43280; AAC43598.1; -
CC PR; JC6040; JC6040.
CC StyGene; SG10609; csgB.
CC Fimbrin; Signal; Complete proteome.
CC SIGNAL 1 21 POTENTIAL.
CC FT CHAIN 22 151 MINOR CURLIN SUBUNIT.
CC SO SEQUENCE 151 AA; 16182 MW; CQFC543086DD361D CRC64;
CC -----
CC Query Match 13.7%; Score 106.5; DB 1; Length 151;
CC Best Local Similarity 30.7%; Pred. No. 0.015;
CC Matches 35; Conservative 17; Mismatches 43; Indels 19; Gaps 5;
CC -----
CC Oy 55 MAHALOSDARKSE-----TTTOSGYGNGADYVG-GADNSF-----IELTONGFR 98
CC Db 18 IATNYDYLARSEYNPAVNELSKSFQAAIIIGVGTDSARVRQBSKLSVISQSGN 77
CC 99 NNATIDQNAKNSDIT-VQGYGGNNALVNTQTSDDSVMYROYFGNNATANQY 151
CC Db 78 NRAVDQ--AGNTYFAIIEQTGNANDSISQSAVGNBAAIIIGKSGKNANTQY 129
CC -----
CC RESULT 7
CC YK03_CABEL
CC AC P34251;
CC ID_YK03_CABEL STANDARD; PRT; 401 AA.
CC DT 01-FEB-1994 (Rel. 28, Created)
CC DT 30-MAY-2000 (Rel. 39, Last sequence update)
CC DT 30-MAY-2000 (Rel. 39, Last annotation update)
CC DE Hypothetical 43.5 kDa protein C05B5.3 in chromosome III.
CC GN C05B5.3.
CC OS Caenorhabditis elegans.
CC OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida; Rhabditoidea;
CC OC Rhabditiidae; Pelodierinae; Caenorhabditis.
CC OX NCBI_Taxid=6239;
CC RN [1]
CC RC SEQUENCE FROM N.A.
CC RC STRAIN=Bristol N2;
CC RA Mortimore B.U.;
CC RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
CC RP REVISIONS.
CC RP Durbin R.;
CC RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; Z32679; CAA83596.1; -
CC DR PIR; C88571; C88571.
CC DR WormPep; C05B5.3; CB17369.
CC KW Hypothetical protein.
CC SO SEQUENCE 401 AA; 43472 MW; C99DE259EF6C8B55 CRC64;
CC -----
CC Query Match 11.8%; Score 92; DB 1; Length 401;
CC Best Local Similarity 23.2%; Pred. No. 0.0;
CC Matches 46; Conservative 19; Mismatches 71; Indels 62; Gaps 7;

```

QY 3 LKVAFAIVSGSALAGVFPWGG-----CGNHNGNSGPDYDLVT 48
 DB 30 LQQRRAVAGCSSGCVATIPKSGFPMADMIAGIOTBRQNONNSNPPODDPT 89
 QY 49 RVTHE-----MAHALQSDARKSETTI---TQSG--YENGADVQGA 85
 DB 90 SQSTQOINGNVGSSSSNQCPVIYIARAGSDKRNSEVTITPTPNGFNFNGQGOQ 149
 QY 86 D-----NSTIELTONGFNNATIDQNNAKNSDITVQOYGNNALVNOTASD 132
 DB 150 NTGSSSGFNNQNONNNQNNQNNQNNQNNQNNQNNQNNQNNQNNQNNQ 197
 QY 133 SSVMVROVFGNNATANO 150
 DB 198 NVGVNPTSGFNSQTSNQ 215

RESULT 8
 OMPB_RICUA STANDARD; PRT; 1656 AA.

AC 006653;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Outer membrane protein B precursor (168 kDa surface-layer protein)
 DE (Surface protein antigen) (Cell surface antigen 5) (Scas) (rOMP B)
 DE (OMP B) [Contains: 120 kDa surface-exposed protein (Surface protein antigen) (120 kDa outer membrane protein ompb); 32 kDa beta peptide].
 GN OMPB.
 OS Rickettsia japonica.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 OX NCBI_Taxid=35790;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=YH;
 RA Uchiyama T.;
 RT "Sequencing of the gene encoding the protein rOMP B of Rickettsia japonica."
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL VIRULANCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (By similarity).
 CC -1- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-layer with hexagonal symmetry.
 CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; AB003681; BAA20138.1; -
 DR InterPro; IPR006315; Autotransporter.
 DR InterPro; IPR005546; Autotransporter.
 DR Pfam; PF03797; Autotransporter; 1.
 DR TIGRFAMs; TIGR01414; autotrans_bar1; 2.
 KW Antigen; s-layer; Cell wall.
 FT CHAIN 1 1338 120 kDa SURFACE-EXPOSED PROTEIN.
 FT CHAIN 1339 1656 32 kDa BETA PEPTIDE.
 FT DOMAIN 528 533 POLY-GLY.
 SQ SSOURCE 1656 AA; 168097 MW; 3132A69CD9D5599F CRC64;

Query March 11.8%; Score 92; DB 1; Length 1656;
 Best Local Similarity 26.2%; Pred. No. 4;
 Matches 45; Conservative 19; Mismatches 52; Indels 56; Gaps 9;

QY 6 VAAFAIVSGSALAGVFPWGGCGNHNGNSGPDYDLVTRVTHEMAHALQSDARK 65
 DB 509 VLAAGAITLDSGSLTI-----TGDINGGGGAA---LQSTI-----LANDARK 547
 QY 66 SETTITQSG-----YENGADVQAGDNSTIELTONGFNNATID----- 104
 DB 548 ---TLTILGANIISANGGTFINFGANGTILKIST--QNNIIVDCDLAITDGTGVASS 602
 QY 105 QNNAKNSDI---TVGOYGNMALL-----VNOTASDSSVMVROVFGNNAT 147
 DB 603 LINAQTLTISGTIGIIGANNITLIGFNPISSTKLTLGVAVINELVIGNNS 654

RESULT 9
 GP63_LEIDO STANDARD; PRT; 590 AA.

AC P23223;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Leishmanolysin precursor (EC 3.4.24.36) (Cell surface protease)
 DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface endopeptidase).
 GN GP63.
 OS Leishmania donovani.
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 OX NCBI_Taxid=5661;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LV9;
 RA MEDLINE=92107220; PubMed=1762629;
 RA Webb J.R., Burton L.V., McMaster R.W.;
 RT "Heterogeneity of the genes encoding the major surface glycoprotein of Leishmania donovani".
 RL Mol. Biochem. Parasitol. 48:173-184(1991).
 CC -1- FUNCTION: Has an integral role during the infection of macrophages in the mammalian host.
 CC -1- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and P1' and basic residues at P2 and P3'. A model nonapeptide is cleaved at -Ala-Tyr-Lys-Lys--.
 CC -1- CORFACTOR: Binds 1 zinc ion per subunit (By similarity).
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
 CC -1- SIMILARITY: Belongs to peptidase family M8.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; M60048; AAA29244.1; -
 DR HSSP; P08148; 1ML.
 DR MEROPS; M08.001; -
 DR InterPro; IPR006025; Pept. M Zn BS.
 DR InterPro; IPR001577; Peptidase_M8.
 DR Pfam; PF01457; Peptidase_M8; 1.
 DR PRINTS; PR00782; LSHMANOLYSIN.
 DR PROSITE; PS00142; ZINC_PROTEASE; 1.
 KW Hydrolyase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;
 KW Zymogen; Signal; Cell adhesion; GPI-anchor; Lipoprotein.
 FT SIGNAL 1 39 POTENTIAL.
 FT PROPEP 40 87 ACTIVATION PEPTIDE.
 FT CHAIN 88 565 LEISHMANOLYSIN.
 FT PROPEP 566 590 REMOVED IN MATURE FORM (BY SIMILARITY).
 FT METAL 251 251 ZINC (CATALYTIC) (BY SIMILARITY).
 FT ACT SITE 252 252 BY SIMILARITY.
 FT METAL 255 255 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 321 321 ZINC (CATALYTIC) (BY SIMILARITY).
 FT DISULFID 112 129 BY SIMILARITY.
 FT DISULFID 178 217 BY SIMILARITY.
 FT DISULFID 301 373 BY SIMILARITY.

FT DISULFID 360 443 BY SIMILARITY.
 FT DISULFID 393 412 BY SIMILARITY.
 FT DISULFID 402 477 BY SIMILARITY.
 FT DISULFID 454 498 BY SIMILARITY.
 FT DISULFID 503 553 BY SIMILARITY.
 FT DISULFID 523 546 BY SIMILARITY.
 FT CARBOHYD 287 565 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT LIPID 565 GPI-anchor amidated asparagine (BY similarity).
 SQ SEQUENCE 590 AA; 62950 MW; 0FB315D29659F58 CRC64;
 Query Match 11.7%; Score 91.5; DB 1; Length 590;
 Best Local Similarity 30.5%; Pred. No. 1.4;
 Matches 43; Conservative 12; Mismatches 39; Indels 47; Gaps 9;
 QY 43 YVOLVTRVTHMAHALQ-----SDAR-----KSETTITQS--GYG 76
 DB 241 YVOLVTRVTHMAHALGFSVFFRDARLLESISNVRKDFDPVYNSTAVAKAEQYQ 300
 QY 77 NGA-----DVGQADNSTIELTONGFRNNATTIDQNAKNSDITVGYGNNALVNO 128
 DB 301 CGTLEYLEMEDQCGAGSAGSHICM-----RNAQ--DELMAPASD--AGYYSALTMALFQD 351
 QY 129 TA---SDSSVMVRQVGFQNN 146
 DB 352 LGFYQADFS-KAEEMFWGRNA 371
 RESULT 10
 PER_DROMI STANDARD; PRT; 1093 AA.
 ID_003297; Q18421; Q18422; P91721; P91722;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Period circadian protein (Fragment).
 GN PER.
 OS Drosophila willistoni (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OC NCBI_TaxId=7260;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=Various strains;
 RX MEDLINE=97357421; PubMed=9214747;
 RA Gleason J.M., Powell J.R.;
 RT "Interspecific and intraspecific comparisons of the period locus in
 RT the Drosophila willistoni sibling species."; Mol. Biol. Evol. 14:741-753(1997).
 RN (2)
 RP SEQUENCE OF 579-646 FROM N.A.
 RX MEDLINE=93196482; PubMed=8450754;
 RA Pelicoto A.A., Camposan S., Costa R.H., Kyriacou C.P.;
 RT "Molecular evolution of a repetitive region within the per gene of
 RT Drosophila."; Mol. Biol. Evol. 10:127-139(1993).
 CC -1- FUNCTION: Essential for biological clock functions. Determines the
 CC period length of circadian and ultradian rhythms; an increase in
 CC PER dosage leads to shortened circadian rhythms and a decrease
 CC leads to lengthened circadian rhythms. Essential for the circadian
 CC rhythmicity of locomotor activity, eclosion behavior, and for the
 CC rhythmic component of the male courtship song that originates in
 CC the thoracic nervous system. The biological cycle depends on the
 CC rhythmic formation and nuclear localization of the TIM-PER
 CC complex. Light induces the degradation of TIM, which promotes
 CC elimination of PER. Nuclear activity of the heterodimer
 CC coordinately regulates PER and TIM transcription through a
 CC negative feedback loop. Behaves as a negative element in circadian
 CC transcriptional loop. Does not appear to bind DNA, suggesting
 CC indirect transcriptional inhibition (By similarity).
 CC -1- SUBUNIT: Forms heterodimer with timeless (TIM); the complex then
 CC translocates into the nucleus (By similarity).

CC -1- SUBCELLULAR LOCATION: Nuclear at specific periods of the day.
 CC First accumulates in the perinuclear region about one hour before
 CC translocation into the nucleus. Interaction with TIM is required
 CC for nuclear localization (By similarity).
 CC -1- PTM: PHOSPHORYLATED WITH A CIRCADIAN RHYTHMICITY, PROBABLY BY THE
 CC DOUBLE-TIME PROTEIN (DBT). PHOSPHORYLATION COULD BE IMPLICATED IN
 CC THE STABILITY OF PER MONOMER AND IN THE FORMATION OF HETERODIMER
 CC PER-TIM (By similarity).
 CC -1- MISCELLANEOUS: THE SEQUENCE SHOWN IS THAT OF STRAIN ATLIXCO.
 CC -1- SIMILARITY: Contains 2 PAS (PER-ARNT-SIM) dimerization domains.
 CC -1- SIMILARITY: Contains 1 PAS-associated C-terminal (PAC) domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U51055; AAB41360.1; -
 CC EMBL; U51056; AAB41361.1; -
 CC EMBL; U51057; AAB41362.1; -
 CC EMBL; U51058; AAB41363.1; -
 CC EMBL; U51059; AAB41364.1; -
 CC EMBL; U51060; AAB41365.1; -
 CC EMBL; U51061; AAB41366.1; -
 CC EMBL; U51062; AAB41367.1; -
 CC EMBL; U51063; AAB41368.1; -
 CC EMBL; U51064; AAB41369.1; -
 CC EMBL; U51065; AAB41370.1; -
 CC EMBL; U51066; AAB41371.1; -
 CC EMBL; U51067; AAB41372.1; -
 CC EMBL; U51068; AAB41373.1; -
 CC EMBL; U51069; AAB41374.1; -
 CC EMBL; U51070; AAB41375.1; -
 CC EMBL; U51071; AAB41376.1; -
 CC EMBL; U51072; AAB41377.1; -
 CC EMBL; U51073; AAB41378.1; -
 CC EMBL; U51074; AAB41379.1; -
 CC EMBL; U51075; AAB41380.1; -
 CC EMBL; U51076; AAB41381.1; -
 CC EMBL; U51077; AAB41382.1; -
 CC EMBL; U51078; AAB41383.1; -
 CC EMBL; U51079; AAB41384.1; -
 CC EMBL; U51080; AAB41385.1; -
 CC EMBL; U51081; AAB41386.1; -
 CC EMBL; U51082; AAB41387.1; -
 CC EMBL; U51083; AAB41388.1; -
 CC EMBL; U51084; AAB41389.1; -
 CC EMBL; U51085; AAB41390.1; -
 CC EMBL; U51086; AAB41391.1; -
 CC EMBL; U51087; AAB41392.1; -
 CC EMBL; U51088; AAB41393.1; -
 CC EMBL; U51089; AAB41394.1; -
 CC EMBL; U51090; AAB41395.1; -
 CC EMBL; U51091; AAB41396.1; -
 CC EMBL; U51092; AAB41397.1; -
 CC EMBL; U51093; AAB41398.1; -
 CC EMBL; U51094; AAB41399.1; -
 CC EMBL; U51095; AAB41400.1; -
 CC EMBL; U51096; AAB41401.1; -
 CC EMBL; U51097; AAB41402.1; -
 CC EMBL; U51098; AAB41403.1; -
 CC EMBL; U51099; AAB41404.1; -
 CC EMBL; U51100; AAB41405.1; -
 CC EMBL; U51101; AAB41406.1; -
 CC EMBL; U51102; AAB41407.1; -
 CC EMBL; U51103; AAB41408.1; -
 CC EMBL; U51104; AAB41409.1; -
 CC EMBL; U51105; AAB41410.1; -
 CC EMBL; U51106; AAB41411.1; -
 CC EMBL; U51107; AAB41412.1; -
 CC EMBL; U51108; AAB41413.1; -
 CC EMBL; U51109; AAB41414.1; -
 CC EMBL; U51110; AAB41415.1; -
 CC EMBL; U51111; AAB41416.1; -
 CC EMBL; U51112; AAB41417.1; -
 CC EMBL; U51113; AAB41418.1; -
 CC EMBL; U51114; AAB41419.1; -
 CC EMBL; U51115; AAB41420.1; -
 CC EMBL; U51116; AAB41421.1; -
 CC EMBL; U51117; AAB41422.1; -
 CC EMBL; U51118; AAB41423.1; -
 CC EMBL; U51119; AAB41424.1; -
 CC EMBL; U51120; AAB41425.1; -
 CC EMBL; U51121; AAB41426.1; -
 CC EMBL; U51122; AAB41427.1; -
 CC EMBL; U51123; AAB41428.1; -
 CC EMBL; U51124; AAB41429.1; -
 CC EMBL; U51125; AAB41430.1; -
 CC EMBL; U51126; AAB41431.1; -
 CC EMBL; U51127; AAB41432.1; -
 CC EMBL; U51128; AAB41433.1; -
 CC EMBL; U51129; AAB41434.1; -
 CC EMBL; U51130; AAB41435.1; -
 CC EMBL; U51131; AAB41436.1; -
 CC EMBL; U51132; AAB41437.1; -
 CC EMBL; U51133; AAB41438.1; -
 CC EMBL; U51134; AAB41439.1; -
 CC EMBL; U51135; AAB41440.1; -
 CC EMBL; U51136; AAB41441.1; -
 CC EMBL; U51137; AAB41442.1; -
 CC EMBL; U51138; AAB41443.1; -
 CC EMBL; U51139; AAB41444.1; -
 CC EMBL; U51140; AAB41445.1; -
 CC EMBL; U51141; AAB41446.1; -
 CC EMBL; U51142; AAB41447.1; -
 CC EMBL; U51143; AAB41448.1; -
 CC EMBL; U51144; AAB41449.1; -
 CC EMBL; U51145; AAB41450.1; -
 CC EMBL; U51146; AAB41451.1; -
 CC EMBL; U51147; AAB41452.1; -
 CC EMBL; U51148; AAB41453.1; -
 CC EMBL; U51149; AAB41454.1; -
 CC EMBL; U51150; AAB41455.1; -
 CC EMBL; U51151; AAB41456.1; -
 CC EMBL; U51152; AAB41457.1; -
 CC EMBL; U51153; AAB41458.1; -
 CC EMBL; U51154; AAB41459.1; -
 CC EMBL; U51155; AAB41460.1; -
 CC EMBL; U51156; AAB41461.1; -
 CC EMBL; U51157; AAB41462.1; -
 CC EMBL; U51158; AAB41463.1; -
 CC EMBL; U51159; AAB41464.1; -
 CC EMBL; U51160; AAB41465.1; -
 CC EMBL; U51161; AAB41466.1; -
 CC EMBL; U51162; AAB41467.1; -
 CC EMBL; U51163; AAB41468.1; -
 CC EMBL; U51164; AAB41469.1; -
 CC EMBL; U51165; AAB41470.1; -
 CC EMBL; U51166; AAB41471.1; -
 CC EMBL; U51167; AAB41472.1; -
 CC EMBL; U51168; AAB41473.1; -
 CC EMBL; U51169; AAB41474.1; -
 CC EMBL; U51170; AAB41475.1; -
 CC EMBL; U51171; AAB41476.1; -
 CC EMBL; U51172; AAB41477.1; -
 CC EMBL; U51173; AAB41478.1; -
 CC EMBL; U51174; AAB41479.1; -
 CC EMBL; U51175; AAB41480.1; -
 CC EMBL; U51176; AAB41481.1; -
 CC EMBL; U51177; AAB41482.1; -
 CC EMBL; U51178; AAB41483.1; -
 CC EMBL; U51179; AAB41484.1; -
 CC EMBL; U51180; AAB41485.1; -
 CC EMBL; U51181; AAB41486.1; -
 CC EMBL; U51182; AAB41487.1; -
 CC EMBL; U51183; AAB41488.1; -
 CC EMBL; U51184; AAB41489.1; -
 CC EMBL; U51185; AAB41490.1; -
 CC EMBL; U51186; AAB41491.1; -
 CC EMBL; U51187; AAB41492.1; -
 CC EMBL; U51188; AAB41493.1; -
 CC EMBL; U51189; AAB41494.1; -
 CC EMBL; U51190; AAB41495.1; -
 CC EMBL; U51191; AAB41496.1; -
 CC EMBL; U51192; AAB41497.1; -
 CC EMBL; U51193; AAB41498.1; -
 CC EMBL; U51194; AAB41499.1; -
 CC EMBL; U51195; AAB41500.1; -
 CC EMBL; U51196; AAB41501.1; -
 CC EMBL; U51197; AAB41502.1; -
 CC EMBL; U51198; AAB41503.1; -
 CC EMBL; U51199; AAB41504.1; -
 CC EMBL; U51200; AAB41505.1; -
 CC EMBL; U51201; AAB41506.1; -
 CC EMBL; U51202; AAB41507.1; -
 CC EMBL; U51203; AAB41508.1; -
 CC EMBL; U51204; AAB41509.1; -
 CC EMBL; U51205; AAB41510.1; -
 CC EMBL; U51206; AAB41511.1; -
 CC EMBL; U51207; AAB41512.1; -
 CC EMBL; U51208; AAB41513.1; -
 CC EMBL; U51209; AAB41514.1; -
 CC EMBL; U51210; AAB41515.1; -
 CC EMBL; U51211; AAB41516.1; -
 CC EMBL; U51212; AAB41517.1; -
 CC EMBL; U51213; AAB41518.1; -
 CC EMBL; U51214; AAB41519.1; -
 CC EMBL; U51215; AAB41520.1; -
 CC EMBL; U51216; AAB41521.1; -
 CC EMBL; U51217; AAB41522.1; -
 CC EMBL; U51218; AAB41523.1; -
 CC EMBL; U51219; AAB41524.1; -
 CC EMBL; U51220; AAB41525.1; -
 CC EMBL; U51221; AAB41526.1; -
 CC EMBL; U51222; AAB41527.1; -
 CC EMBL; U51223; AAB41528.1; -
 CC EMBL; U51224; AAB41529.1; -
 CC EMBL; U51225; AAB41530.1; -
 CC EMBL; U51226; AAB41531.1; -
 CC EMBL; U51227; AAB41532.1; -
 CC EMBL; U51228; AAB41533.1; -
 CC EMBL; U51229; AAB41534.1; -
 CC EMBL; U51230; AAB41535.1; -
 CC EMBL; U51231; AAB41536.1; -
 CC EMBL; U51232; AAB41537.1; -
 CC EMBL; U51233; AAB41538.1; -
 CC EMBL; U51234; AAB41539.1; -
 CC EMBL; U51235; AAB41540.1; -
 CC EMBL; U51236; AAB41541.1; -
 CC EMBL; U51237; AAB41542.1; -
 CC EMBL; U51238; AAB41543.1; -
 CC EMBL; U51239; AAB41544.1; -
 CC EMBL; U51240; AAB41545.1; -
 CC EMBL; U51241; AAB41546.1; -
 CC EMBL; U51242; AAB41547.1; -
 CC EMBL; U51243; AAB41548.1; -
 CC EMBL; U51244; AAB41549.1; -
 CC EMBL; U51245; AAB41550.1; -
 CC EMBL; U51246; AAB41551.1; -
 CC EMBL; U51247; AAB41552.1; -
 CC EMBL; U51248; AAB41553.1; -
 CC EMBL; U51249; AAB41554.1; -
 CC EMBL; U51250; AAB41555.1; -
 CC EMBL; U51251; AAB41556.1; -
 CC EMBL; U51252; AAB41557.1; -
 CC EMBL; U51253; AAB41558.1; -
 CC EMBL; U51254; AAB41559.1; -
 CC EMBL; U51255; AAB41560.1; -
 CC EMBL; U51256; AAB41561.1; -
 CC EMBL; U51257; AAB41562.1; -
 CC EMBL; U51258; AAB41563.1; -
 CC EMBL; U51259; AAB41564.1; -
 CC EMBL; U51260; AAB41565.1; -
 CC EMBL; U51261; AAB41566.1; -
 CC EMBL; U51262; AAB41567.1; -
 CC EMBL; U51263; AAB41568.1; -
 CC EMBL; U51264; AAB41569.1; -
 CC EMBL; U51265; AAB41570.1; -
 CC EMBL; U51266; AAB41571.1; -
 CC EMBL; U51267; AAB41572.1; -
 CC EMBL; U51268; AAB41573.1; -
 CC EMBL; U51269; AAB41574.1; -
 CC EMBL; U51270; AAB41575.1; -
 CC EMBL; U51271; AAB41576.1; -
 CC EMBL; U51272; AAB41577.1; -
 CC EMBL; U51273; AAB41578.1; -
 CC EMBL; U51274; AAB41579.1; -
 CC EMBL; U51275; AAB41580.1; -
 CC EMBL; U51276; AAB41581.1; -
 CC EMBL; U51277; AAB41582.1; -
 CC EMBL; U51278; AAB41583.1; -
 CC EMBL; U51279; AAB41584.1; -
 CC EMBL; U51280; AAB41585.1; -
 CC EMBL; U51281; AAB41586.1; -
 CC EMBL; U51282; AAB41587.1; -
 CC EMBL; U51283; AAB41588.1; -
 CC EMBL; U51284; AAB41589.1; -
 CC EMBL; U51285; AAB41590.1; -
 CC EMBL; U51286; AAB41591.1; -
 CC EMBL; U51287; AAB41592.1; -
 CC EMBL; U51288; AAB41593.1; -
 CC EMBL; U51289; AAB41594.1; -
 CC EMBL; U51290; AAB41595.1; -
 CC EMBL; U51291; AAB41596.1; -
 CC EMBL; U51292; AAB41597.1; -
 CC EMBL; U51293; AAB41598.1; -
 CC EMBL; U51294; AAB41599.1; -
 CC EMBL; U51295; AAB41600.1; -
 CC EMBL; U51296; AAB41601.1; -
 CC EMBL; U51297; AAB41602.1; -
 CC EMBL; U51298; AAB41603.1; -
 CC EMBL; U51299; AAB41604.1; -
 CC EMBL; U51300; AAB41605.1; -
 CC EMBL; U51301; AAB41606.1; -
 CC EMBL; U51302; AAB41607.1; -
 CC EMBL; U51303; AAB41608.1; -
 CC EMBL; U51304; AAB41609.1; -
 CC EMBL; U51305; AAB41610.1; -
 CC EMBL; U51306; AAB41611.1; -
 CC EMBL; U51307; AAB41612.1; -
 CC EMBL; U51308; AAB41613.1; -
 CC EMBL; U51309; AAB41614.1; -
 CC EMBL; U51310; AAB41615.1; -
 CC EMBL; U51311; AAB41616.1; -
 CC EMBL; U51312; AAB41617.1; -
 CC EMBL; U51313; AAB41618.1; -
 CC EMBL; U51314; AAB41619.1; -
 CC EMBL; U51315; AAB41620.1; -
 CC EMBL; U51316; AAB41621.1; -
 CC EMBL; U51317; AAB41622.1; -
 CC EMBL; U51318; AAB41623.1; -
 CC EMBL; U51319; AAB41624.1; -
 CC EMBL; U51320; AAB41625.1; -
 CC EMBL; U51321; AAB41626.1; -
 CC EMBL; U51322; AAB41627.1; -
 CC EMBL; U51323; AAB41628.1; -
 CC EMBL; U51324; AAB41629.1; -
 CC EMBL; U51325; AAB41630.1; -
 CC EMBL; U51326; AAB41631.1; -
 CC EMBL; U51327; AAB41632.1; -
 CC EMBL; U51328; AAB41633.1; -
 CC EMBL; U51329; AAB41634.1; -
 CC EMBL; U51330; AAB41635.1; -
 CC EMBL; U51331; AAB41636.1; -
 CC EMBL; U51332; AAB41637.1; -
 CC EMBL; U51333; AAB41638.1; -
 CC EMBL; U51334; AAB41639.1; -
 CC EMBL; U51335; AAB41640.1; -
 CC EMBL; U51336; AAB41641.1; -
 CC EMBL; U51337; AAB41642.1; -
 CC EMBL; U51338; AAB41643.1; -
 CC EMBL; U51339; AAB41644.1; -
 CC EMBL; U51340; AAB41645.1; -
 CC EMBL; U51341; AAB41646.1; -
 CC EMBL; U51342; AAB41647.1; -
 CC EMBL; U51343; AAB41648.1; -
 CC EMBL; U51344; AAB41649.1; -
 CC EMBL; U51345; AAB41650.1; -
 CC EMBL; U51346; AAB41651.1; -
 CC EMBL; U51347; AAB41652.1; -
 CC EMBL; U51348; AAB41653.1; -
 CC EMBL; U51349; AAB41654.1; -
 CC EMBL; U51350; AAB41655.1; -
 CC EMBL; U51351; AAB41656.1; -
 CC EMBL; U51352; AAB41657.1; -
 CC EMBL; U51353; AAB41658.1; -
 CC EMBL; U51354; AAB41659.1; -
 CC EMBL; U51355; AAB41660.1; -
 CC EMBL; U51356; AAB41661.1; -
 CC EMBL; U51357; AAB41662.1; -
 CC EMBL; U51358; AAB41663.1; -
 CC EMBL; U51359; AAB41664.1; -
 CC EMBL; U51360; AAB41665.1; -
 CC EMBL; U51361; AAB41666.1; -
 CC EMBL; U51362; AAB41667.1; -
 CC EMBL; U51363; AAB41668.1; -
 CC EMBL; U51364; AAB41669.1; -
 CC EMBL; U51365; AAB41670.1; -
 CC EMBL; U51366; AAB41671.1; -
 CC EMBL; U51367; AAB41672.1; -
 CC EMBL; U51368; AAB41673.1; -
 CC EMBL; U51369; AAB41674.1; -
 CC EMBL; U51370; AAB41675.1; -
 CC EMBL; U51371; AAB41676.1; -
 CC EMBL; U51372; AAB41677.1; -
 CC EMBL; U51373; AAB41678.1; -
 CC EMBL; U51374; AAB41679.1; -
 CC EMBL; U51375; AAB41680.1; -
 CC EMBL; U51376; AAB41681.1; -
 CC EMBL; U51377; AAB41682.1; -
 CC EMBL; U51378; AAB41683.1; -
 CC EMBL; U51379; AAB41684.1; -
 CC EMBL; U51380; AAB41685.1; -
 CC EMBL; U51381; AAB41686.1; -
 CC EMBL; U51382; AAB41687.1; -
 CC EMBL; U51383; AAB41688.1; -
 CC EMBL; U51384; AAB41689.1; -
 CC EMBL; U51385; AAB41690.1; -
 CC EMBL; U51386; AAB41691.1; -
 CC EMBL; U51387; AAB41692.1; -
 CC EMBL; U51388; AAB41693.1; -
 CC EMBL; U51389; AAB41694.1; -
 CC EMBL; U51390; AAB41695.1; -
 CC EMBL; U51391; AAB41696.1; -
 CC EMBL; U51392; AAB41697.1; -
 CC EMBL; U51393; AAB41698.1; -
 CC EMBL; U51394; AAB41699.1; -
 CC EMBL; U51395; AAB41700.1; -
 CC EMBL; U51396; AAB41701.1; -
 CC EMBL; U51397; AAB41702.1; -
 CC EMBL; U51398; AAB41703.1; -
 CC EMBL; U51399; AAB41704.1; -
 CC EMBL; U51400; AAB41705.1; -
 CC EMBL; U51401; AAB41706.1; -
 CC EMBL; U51402; AAB41707.1; -
 CC EMBL; U51403; AAB41708.1; -
 CC EMBL; U51404; AAB41709.1; -
 CC EMBL; U51405; AAB41710.1; -
 CC EMBL; U51406; AAB41711.1; -
 CC EMBL; U51407; AAB41712.1; -
 CC EMBL; U51408; AAB41713.1; -
 CC EMBL; U51409; AAB41714.1; -
 CC EMBL; U51410; AAB41715.1; -
 CC EMBL; U51411; AAB41716.1; -
 CC EMBL; U51412; AAB41717.1; -
 CC EMBL; U51413; AAB41718.1; -
 CC EMBL; U51414; AAB41719.1; -
 CC EMBL; U51415; AAB41720.1; -
 CC EMBL; U51416; AAB41721.1; -
 CC EMBL; U51417; AAB41722.1; -
 CC EMBL; U51418; AAB41723.1; -
 CC EMBL; U51419; AAB41724.1; -
 CC EMBL; U51420; AAB41725.1; -
 CC EMBL; U51421; AAB41726.1; -
 CC EMBL; U51422; AAB41727.1; -
 CC EMBL; U51423; AAB41728.1; -
 CC EMBL; U51424; AAB41729.1; -
 CC EMBL; U51425; AAB41730.1; -
 CC EMBL; U51426; AAB4173

FT VARIANT 734 734 MISSING (IN STRAINS MANAUS 4, PORTO
 FT VARIANT 747 747 ALGEBRE 1 AND PORTO ALBRE 2).
 FT VARIANT 764 766 S -> A (IN STRAINS GUADELOUPE AND GUANA).
 FT VARIANT 886 886 MISSING (IN STRAIN MANAUS 3).
 FT NON_TER 1093 1093 A -> T (IN STRAIN 0811.4).
 SQ SEQUENCE 1093 AA; 115896 MW; AB6DE050267EC187 CRC64;
 Query Match 11.7%; Score 91; DB 1; Length 1093;
 Best Local Similarity 26.1%; Pred. No. 3.1;
 Matches 24; Conservative 8; Mismatches 46; Indels 14; Gaps 2;
 QY 27 GGGGNNHGGGSSGPDYDLYTRVTHVEMAHALQSPARSETTITQSGVNGADVGGQAD 86
 DB 719 GGGGGGGGGGGGGGGGGLPLFL-----DVTHTSSSQNGKPTGVAAAGAGGGGGGG-- 770
 QY 87 NSTIELTONGFRNNAITIDDMNAKNSDITVGY 118
 DB 771 -----SCSGLGNGNGVSGNGNNSQPSITNY 796
 RESULT 11
 GP63 LEIMA STANDARD; PRT; 602 AA.
 AC P08148; P15906; (Rel. 08, Created)
 DT 01-AUG-1988 (Rel. 14, Last sequence update)
 DT 01-APR-1990 (Rel. 42, Last annotation update)
 DE Leishmanolysin precursor (EC 3.4.24.36) (Cell surface protease)
 DE (Major surface glycoprotein) (gp63 protein) (Promastigote surface
 endopeptidase).
 GN GP63.
 OS Leishmania major.
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 OX NCBI_TaxID=5664;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 101-123.
 RX MEDLINE=88154764; PubMed=3346625;
 RA Burton L.L., McMaster W.R.;
 RT "Molecular cloning of the major surface antigen of Leishmania.",
 RL J. Exp. Med. 167:724-729 (1988).
 RN [2]
 RP REVISIONS.
 RA Burton L.L., McMaster W.R.;
 RL J. Exp. Med. 171:589-589 (1990).
 RN [3]
 RP GPI-ANCHOR.
 RX MEDLINE=91099116; PubMed=2145267;
 RA Schneider P., Ferguson M.A.J., McConville M.J., Mehler A.,
 RT Homans S.W., Bordier C.;
 RT "Structure of the glycosyl-phosphatidylinositol membrane anchor of
 the Leishmania major promastigote surface protease.",
 RL J. Biol. Chem. 265:16955-16964 (1990).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).
 RX MEDLINE=95406217; PubMed=7675788;
 RA Schlagenhauf E., Egges R., Metcalf P.;
 RT "Crystallization and preliminary X-ray diffraction studies of
 Leishmanolysin, the major surface metalloproteinase from Leishmania
 major.",
 RL Proteins 22:58-66 (1995).
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY (1.86 ANGSTROMS).
 RX MEDLINE=98416698; PubMed=9739094;
 RA Schlagenhauf E., Egges R., Metcalf P.;
 RT "The crystal structure of the Leishmania major surface proteinase
 Leishmanolysin.",
 RL Structure 6:1035-1046 (1998).
 CC -!- FUNCTION: Has an integral role during the infection of macrophages
 CC in the mammalian host.
 CC -!- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and
 CC P2 and basic residues at P3 and P3'. A model nonapeptide is
 CC cleaved at -Ala-Tyr-Lys-Lys-.

CC -!- COFACTOR: Binds 1 zinc ion per subunit.
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
 CC -!- PTM: THE PHOSPHATIDYLINOSITOL MOIETY OF THE GPI-ANCHOR CONTAINS A
 CC FULLY SATURATED, UNBRANCHED 1-O-ALKYL CHAIN (MAINLY C24:0) AND A
 CC MIXTURE OF FULLY SATURATED UNBRANCHED 2-O-ACYL CHAINS (C12:0,
 CC C14:0, C16:0, AND C18:0).
 CC -!- SIMILARITY: Belongs to peptidase family M8.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; Y00647; CAA68673.1; -
 CC PIR; P10221; P10221.
 CC PDB; 1LMJ; 17-SEP-97.
 CC MEROPS; M08.001; -
 CC InterPro; IPR006025; Pept_M_Zn_BS.
 CC InterPro; IPR001577; Peptidase_M8.
 CC Pfam; PF01457; Peptidase_M8; 1.
 CC PRINTS; PR00782; LSHMANOLYSIN.
 CC PROSITE; PS00142; ZINC_PROTEASE; 1.
 CC Hydroxylase, Metalloprotease; Glycoprotein; Metal-binding; Zinc;
 CC KMW Zymogen; Signal; Cell adhesion; GPI-anchor; 3D-structure; Lipoprotein.
 CC FT SIGNAL 1 39
 CC FT PROPEP 40 100
 CC FT CHAIN 101 577
 CC FT PROPEP 578 602
 CC FT METAL 264 264
 CC FT ACT_SITE 265 265
 CC FT METAL 268 268
 CC FT METAL 334 334
 CC FT DISULFID 125 142
 CC FT DISULFID 191 230
 CC FT DISULFID 314 386
 CC FT DISULFID 393 455
 CC FT DISULFID 406 425
 CC FT DISULFID 415 489
 CC FT DISULFID 466 510
 CC FT DISULFID 515 565
 CC FT DISULFID 535 558
 CC FT CARBOHYD 300 300
 CC FT CARBOHYD 407 407
 CC FT LIPID 577 577
 CC FT STRAND 101 102
 CC FT STRAND 107 108
 CC FT STRAND 111 114
 CC FT HELIX 116 119
 CC FT TURN 121 122
 CC FT TURN 128 129
 CC FT STRAND 131 133
 CC FT STRAND 139 141
 CC FT HELIX 144 146
 CC FT HELIX 150 158
 CC FT TURN 159 159
 CC FT HELIX 160 169
 CC FT TURN 170 171
 CC FT STRAND 172 174
 CC FT STRAND 177 178
 CC FT STRAND 180 181
 CC FT TURN 189 190
 CC FT HELIX 191 193
 CC FT HELIX 198 198
 CC FT TURN 203 203
 CC FT STRAND 205 206
 CC FT STRAND 210 215
 CC FT TURN 221 222
 CC FT STRAND 226 232
 CC FT TURN 234 235
 CC FT STRAND 238 244

N-LINKED (GLCNAC. . .) (POTENTIAL).
 N-ANCHOR (GLCNAC. . .) (POTENTIAL).
 GPI-ANCHOR amidated asparagine.

```

FT HELIX 247 249
FT HELIX 256 269
FT TURN 270 271
FT HELIX 274 279
FT TURN 280 281
FT STRAND 283 286
FT HELIX 289 291
FT STRAND 296 299
FT HELIX 302 312
FT TURN 313 313
FT TURN 315 316
FT STRAND 320 322
FT TURN 328 332
FT STRAND 335 335
FT TURN 337 339
FT TURN 341 342
FT STRAND 343 343
FT TURN 344 345
FT STRAND 353 353
FT HELIX 356 364
FT TURN 365 366
FT STRAND 369 370
FT HELIX 372 374
FT TURN 380 383
FT HELIX 386 390
FT STRAND 394 395
FT TURN 396 397
FT STRAND 398 399
FT TURN 402 404
FT STRAND 413 414
FT TURN 417 418
FT STRAND 421 425
FT STRAND 428 429
FT HELIX 435 437
FT TURN 443 444
FT STRAND 445 446
FT TURN 450 454
FT STRAND 458 465
FT TURN 466 467
FT HELIX 470 472
FT TURN 475 477
FT HELIX 480 480
FT TURN 485 486
FT STRAND 487 494
FT STRAND 496 496
FT STRAND 506 516
FT TURN 517 520
FT STRAND 521 525
FT TURN 527 528
FT STRAND 533 534
FT TURN 537 538
FT STRAND 540 542
FT HELIX 543 545
FT TURN 546 546
FT STRAND 550 553
FT TURN 552 553
FT STRAND 555 557
FT HELIX 561 565
FT TURN 566 567
FT HELIX 569 572
FT TURN 573 573
SQ SEQUENCE 602 AA; 63953 MW; 982EF3245D87C43E CRC64;

```

```

Query Match 11.5%; Score 90; DB 1; Length 602;
Best Local Similarity 33.3%; Pred. No. 1.9;
Matches 28; Conservative 6; Mismatches 16; Indels 34; Gaps 3;

```

```

QY 43 YQQLVTRVYTHHEAAALQ-----SDAR-----KSETTTTQSGYNG 78
DB 254 YQQLVTRVYTHHEAAALGSGPFEDARIVANVNVKGNKDVIVINSSTAFAKAKAEQYG 313
QY 79 ADV-----GCGANDSTIEL 92

```

```

DB 314 CDTLEYEVEDQGAGSAGSHIM 337
RESULT 12
GLR1_HUMAN STANDARD; PRT; 906 AA.
AC P42261;
ID GLR1_HUMAN
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Glutamate receptor 1 precursor (GLUR-1) (GLUR-A) (GLUR-K1) (Glutamate
DE receptor ionotropic, AMPA 1).
GN GR1A1 OR GLUR1 OR GLUR1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92159065; PubMed=1311100;
RA Sun W., Ferrer-Montiel A.V., Schinder A.F., McPherson J.P.,
RA Evans G.A., Montal M.;
RT "Molecular cloning, chromosomal mapping, and functional expression of
RT human brain glutamate receptors."
RL Proc. Natl. Acad. Sci. U.S.A. 89:1443-1447(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX TISSUE=Hippocampus;
RC MEDLINE=92329975; PubMed=130959;
RA Potier M.C., Spiliantini M.G., Carter N.P.;
RT "The human glutamate receptor cDNA GLUR1: cloning, sequencing,
RT expression and localization to chromosome 5."
RL DNA Seq. 2:211-218(1992).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=91352027; PubMed=1652753;
RA Puckett C., Gomez C.M., Korenberg J.R., Tung H., Meier T.J.,
RA Chen X.N., Hood L.E.;
RT "Molecular cloning and chromosomal localization of one of the human
RT glutamate receptor genes."
RL Proc. Natl. Acad. Sci. U.S.A. 88:7557-7561(1991).
CC -!- FUNCTION: L-glutamate acts as an excitatory neurotransmitter at
CC many synapses in the central nervous system. The postsynaptic
CC actions of Glu are mediated by a variety of receptors that are
CC named according to their selective agonists.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Flp;
CC IsoId=P42261-1; Sequence=Displayed;
CC Name=Flp;
CC IsoId=P42261-2; Sequence=VSP_000092, VSP_000093, VSP_000094,
CC VSP_000095, VSP_000096;
CC -!- TISSUE SPECIFICITY: Widely expressed in brain.
CC -!- MISCELLANEOUS: This receptor binds AMPA (quisqualate) > glutamate >
CC kainate.
CC -!- SIMILARITY: Belongs to the ligand-gated ionic channel family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib.ch).
CC -----
CC EMBL; M81886; AA58395.1; -
CC EMBL; X58633; CA41491.1; -
CC EMBL; M64752; AA58613.1; -
CC PIR; A40222; A40222.
CC PIR; S25852; S25852.
CC HSSP; P19491; 1GR2.
CC Genew; HGNC:4571; GR1A1.

```

```

DR MW; 138248;
DR GO; GO:0005986; C:plasma membrane; TAS.
DR GO; GO:0008066; F:glutamate receptor activity; TAS.
DR GO; GO:0015277; F:kainate selective glutamate receptor activity; TAS.
DR GO; GO:0007165; P:signal transduction; TAS.
DR GO; GO:0007268; P:synaptic transmission; TAS.
DR InterPro; IPR001828; ANF_receptor.
DR InterPro; IPR001320; Ion_glu_receptor.
DR InterPro; IPR001622; K-channel_pore.
DR InterPro; IPR001508; NMDA_receptor.
DR InterPro; IPR001311; SAP/glu_receptor.
DR Pfam; PF01094; ANF_receptor; 1.
DR Pfam; PF00060; lig_chan; 1.
DR PRINTS; PR00177; NMDARECEPTOR.
DR SMART; SM00079; PBPE; 1.
DR Receptor; Postsynaptic membrane; Ionic channel; Glycoprotein; Signal;
KM Transmembrane; Alternative splicing.
FT CHAIN 1 18 POTENTIAL.
FT DOMAIN 19 906 GLUTAMATE RECEPTOR 1.
FT TRANSMEM 537 557 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 585 605 POTENTIAL.
FT TRANSMEM 618 638 POTENTIAL.
FT TRANSMEM 806 826 POTENTIAL.
FT VARSPIC 758 758 N -> G (in isoform flp).
FT VARSPIC 768 768 /FtId=VSP 000092.
FT VARSPIC 772 772 N -> S (in isoform flp).
FT VARSPIC 778 778 L -> V (in isoform flp).
FT VARSPIC 788 788 /FtId=VSP_000094.
FT VARSPIC 790 793 N -> S (in isoform flp).
FT VARSPIC 793 793 /FtId=VSP 000095.
FT VARSPIC 793 793 GGCG -> KDSG (in isoform flp).
FT VARSPIC 793 793 /FtId=VSP 000096.
FT CONFLICT 345 345 R -> A (IN REF. 2).
FT CONFLICT 375 375 S -> G (IN REF. 2 AND 3).
FT CONFLICT 863 863 S -> SA (IN REF. 3).
FT CONFLICT 865 867 AGA -> TAP (IN REF. 3).
SO SEQUENCE 906 AA; 101536 MW; 03EALIE0ED10E8EF CRC64;

Query Match 11.5%; Score 90; DB 1; Length 906;
Best Local Similarity 22.6%; Pred. No. 3;
Matches 45; Conservative 26; Mismatches 56; Indels 72; Gaps 10;

QY 6 VAAFAIVVSGSALAGVPMWGGGNNNGSGSPYDOL-----VTRVTH-----M 55
DB 250 VTGQLVAVYDTIPAKIMQW-----KNSDARHTRVDMKRPKTYTSALTYDGVKM 300
QY 56 AAHALQSDARKSETTITQSGYNGADV-----GGGAD-----NSTIELTQ 94
DB 301 AEAFOSLRRQ---RIDISIRGNAGDCLANPAVPMWGGCIDIQBALQVRFGLNGVQFNE 357
QY 95 NGPRNNMT-----IDWNAKNSDI---TYGVGGNNAALVNOT-----ASD 132
DB 358 KGRNTVTLHAVIEKHKDSIRKIIGWNEDKFEVPATDAQGGDNSSVQNTTYVTTLIED 417
QY 133 SSVMAROVGFENNTANQY 151
DB 418 PYWMLKK-----NANQF 429

RESULT 13
OMPB_RICCN STANDARD; PRT; 1655 AA.
AC Q9KKAJ; Q9KKAJ; Q9XC45;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Outer membrane protein B precursor (168 kDa surface-layer protein)
DE (Surface protein antigen) (Cell surface antigen 5) (Scas) (OMPb)
DE (OMP B) [Contains: 120 kDa surface-exposed protein (Surface protein
antigen) (120 kDa outer membrane protein ompb); 32 kDa beta peptide].
GN OMPB OR R1085.

```

```

OS Rickettsia conorii.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsiidae; Rickettsia.
OX NCBI_TaxID=781;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Malish 7;
RX MEDLINE=21442074; PubMed=11557893;
RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
RA Samson D., Roux V., Cosset P., Weissenbach J., Claverie J.-M.,
RA Raoult D.;
RL "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.",
Science 293:2093-2098(2001).
RN [2]
RP SEQUENCE OF 33-1649 FROM N.A.
RC STRAIN=Indian tick typhus, and Malish 7;
RX MEDLINE=20393643; PubMed=10939649;
RA Roux V., Raoult D.;
RT "Phylogenetic analysis of members of the genus Rickettsia using the
RN gene coding the outer-membrane protein ompb (ompb).";
RN Int. J. Syst. Evol. Microbiol. 50:1443-1455(2000).
RP SEQUENCE OF 353-1655 FROM N.A.
RC STRAIN=Malish 7;
RA Stenos J., Walker D.;
RT "The rickettsial outer membrane protein A and B genes of Rickettsia
RN australis, the most divergent rickettsia of the spotted fever group.";
Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (BY
similarity).
CC -1- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR
(BY similarity).
CC -1- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-
layer with hexagonal symmetry (By similarity).
CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAL OMPB/OMP FAMILY.
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AE008659; AAL03623.1; -
DR EMBL; AF123721; AAF34124.1; -
DR EMBL; AF123726; AAF34129.1; -
DR EMBL; AF149110; AAD39533.1; -
DR PIR; E97835; E97835.
DR InterPro; IPR006315; Autotransport.
DR InterPro; IPR005546; Autotransporter.
DR Pfam; PF03797; Autotransporter; 1.
DR TIGRPFAM; TIGR01414; autotrans_barl; 2.
KW Antigen; S-layer; Cell wall; Complete proteome.
FT CHAIN 1 1334 120 KDA SURFACE-EXPOSED PROTEIN.
FT CHAIN 1335 1655 32 KDA BETA PEPTIDE.
FT VARIANT 61 61 P -> A (IN STRAIN INDIAN TICK TYPHUS).
FT VARIANT 75 75 G -> S (IN STRAIN INDIAN TICK TYPHUS).
FT VARIANT 78 78 K -> N (IN STRAIN INDIAN TICK TYPHUS).
FT VARIANT 251 251 V -> A (IN STRAIN INDIAN TICK TYPHUS).
FT VARIANT 413 413 N -> D (IN STRAIN INDIAN TICK TYPHUS).
FT VARIANT 959 959 I -> V (IN STRAIN INDIAN TICK TYPHUS).
FT VARIANT 988 988 A -> T (IN STRAIN INDIAN TICK TYPHUS).
FT VARIANT 1139 1139 R -> L (IN STRAIN INDIAN TICK TYPHUS).
FT CONFLICT 353 354 KD -> GH (IN REF. 3).
FT CONFLICT 354 354 F -> S (IN REF. 3).
FT CONFLICT 776 776 E -> D (IN REF. 3).
FT CONFLICT 1159 1159 G -> S (IN REF. 3).
FT CONFLICT 1177 1177 E -> S (IN REF. 3).
FT CONFLICT 1492 1492 H -> R (IN REF. 3).
SO SEQUENCE 1655 AA; 168342 MW; E49E19377D5FCB7 CRC64;

```



```
CC ----- This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U11383; AAB60216.1; -.
DR EMBL; X59772; CAB36921.1; ALT_SEQ.
DR PIR; A56038; A56038.
DR HSSP; P07248; 2ADR.
DR TRASEPAC; T00669; -.
DR pYBase; FBgn0003028; ovo.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00096; zf-C2H2; 3.
DR SMART; SM00355; Znf_C2H2; 4.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 3.
KM Zinc-finger; Metal-binding; DNA-binding; Repeat; Nuclear protein;
KM Transcription regulation.
FT DOMAIN 62 66 POLY-ALA.
FT DOMAIN 72 77 POLY-GLY.
FT DOMAIN 80 85 POLY-GLY.
FT DOMAIN 98 108 POLY-GLY.
FT DOMAIN 144 152 POLY-HTS.
FT DOMAIN 153 159 POLY-ASN.
FT DOMAIN 336 339 POLY-GLN.
FT DOMAIN 347 353 POLY-GLN.
FT DOMAIN 357 361 POLY-GLN.
FT DOMAIN 410 414 POLY-GLN.
FT DOMAIN 418 422 POLY-GLN.
FT DOMAIN 426 432 POLY-GLN.
FT DOMAIN 445 453 POLY-GLN.
FT DOMAIN 456 459 POLY-GLN.
FT DOMAIN 466 474 POLY-GLN.
FT DOMAIN 497 517 POLY-ALA.
FT DOMAIN 524 529 POLY-SER.
FT DOMAIN 549 558 POLY-ALA.
FT DOMAIN 639 651 POLY-ALA.
FT DOMAIN 717 725 POLY-ALA.
FT DOMAIN 797 802 POLY-GLN.
FT DOMAIN 820 823 POLY-GLN.
FT DOMAIN 826 832 POLY-GLN.
FT ZN_FING 874 896 C2H2-TYPE 1.
FT ZN_FING 902 924 C2H2-TYPE 2.
FT ZN_FING 930 953 C2H2-TYPE 3.
FT ZN_FING 969 992 C2H2-TYPE 4.
FT CONFLICT 647 647 A -> R (IN REF. 2).
SQ SEQUENCE 1028 AA; 110620 MW; D7068B2BC0F6F77 CRC64;

Query Match 11.2%; Score 87.5; DB 1; Length 1028;
Best Local Similarity 25.6%; Pred. No. 5.7;
Matches 40; Conservative 11; Mismatches 58; Indels 47; Gaps 6;

QY 3 LKAAFAAIVVSGSALAGVPPQWGGGNNNGGSSGPDYDLVTRVTHMAHALQSD 62
DB 59 LQNAALAAIYMSAGS-----GGCTGNGGGGASGP----- 89
QY 63 ARKSETTITOSGYNGADVQAGADN-----STIELTONGFRNNATIDQWAKNSDI--- 113
DB 90 ---CGGPSANSGGGGGGGGGNGYINCGVGVPNNSLDGNMLNFAVSNTNESKPFNH 146
QY 114 -TVGQYGNNAALVNOTASDSVVRQVGFQNNATA 148
DB 147 HHHHGHNNNN-----NNNGGQTSMMGHPE-YGANPSA 177
```

Search completed: March 11, 2004, 18:34:56
Job time : 7.3 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:23:54 ; Search time 30.5 Seconds
(without alignments)
1562.074 Million cell updates/sec

Title: US-09-543-407-18

Perfect score: 780
Sequence: 1 MLLKLVAAFAAIVSGSALA.....DSSVMVROYFGNNATANQY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPREMBL_25: *
1: sp_archaea: *
2: sp_bacteria: *
3: sp_fungi: *
4: sp_human: *
5: sp_invertebrate: *
6: sp_mammal: *
7: sp_mhc: *
8: sp_organelle: *
9: sp_phage: *
10: sp_plant: *
11: sp_rodent: *
12: sp_virus: *
13: sp_vertebrate: *
14: sp_unclassified: *
15: sp_virus: *
16: sp_bacteriophage: *
17: sp_archae: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	682	87.4	152	2	033802 salmonella
2	593.5	76.1	150	2	07X243 citrobacter
3	550	70.5	149	2	07X240 citrobacter
4	498.5	63.9	152	16	08CW63 escherichia
5	428.5	54.9	150	2	07X237 enterobacter
6	303	38.8	76	2	054069 salmonella
7	122	15.6	29	2	0953J5 escherichia
8	114	14.6	1748	5	094821 terrahymena
9	113	14.5	502	16	08E1H4 shewanella
10	110.5	14.2	151	2	07X244 citrobacter
11	108	13.8	171	16	089J13 bradyrhizob
12	107.5	13.8	151	16	07UC21 shigella fl
13	107.5	13.8	160	16	08CW64 escherichia
14	107.5	13.8	160	16	083RU7 shigella fl
15	106.5	13.7	1209	16	083CK5 bradyrhizob
16	105.5	13.5	91	2	0953J8 escherichia

17	101.5	13.0	154	16	089J15 bradyrhizob
18	101	12.9	262	5	09VIX5 drosophila
19	100	12.8	179	2	033801 salmonella
20	99	12.7	130	16	089J14 bradyrhizob
21	98	12.6	157	16	088HG0 pseudomonas
22	98	12.6	362	16	08EV64 mycoplasma
23	98	12.6	3659	16	088LNG rhizobium
24	97.5	12.5	151	2	07X238 enterobacter
25	97.5	12.5	152	2	07X241 citrobacter
26	97	12.4	368	16	08EWD6 mycoplasma
27	97	12.4	1422	16	08EPU3 shewanella
28	97	12.4	2035	2	09XC14 salmonella
29	97	12.4	2039	16	082N57 salmonella
30	96.5	12.4	1765	16	07V855 prochlorococcus
31	96	12.3	490	16	08EY9 leptospira
32	95.5	12.2	145	16	08U6N9 agrobacteri
33	95.5	12.2	1613	2	09KKB2 israeli tic
34	95.5	12.2	3501	16	08Y106 raietonia s
35	95.5	12.2	3552	16	08XSD6 raietonia s
36	94.5	12.1	153	16	089J16 bradyrhizob
37	94.5	12.1	348	13	093397 cyprinus ca
38	94	12.1	480	16	089EV2 bradyrhizob
39	93.5	12.0	139	16	08E1H3 shewanella
40	93.5	12.0	287	5	09VIX6 drosophila
41	93.5	12.0	582	16	071868 mycobacteri
42	93.5	12.0	582	16	07TW98 mycobacteri
43	93.5	12.0	1615	2	09KKX8 rickettsia
44	93	11.9	141	16	08U6P1 agrobacteri
45	92	11.8	353	16	08EV92 mycoplasma

ALIGNMENTS

RESULT 1	
033802	PRELIMINARY; PRT; 152 AA.
ID	033802
AC	033802;
DT	01-JAN-1998 (TREMBLrel. 05, Created)
DT	01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE	Agfa protein (Fragment).
GN	Agfa.
OS	Salmonella typhimurium.
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC	Enterobacteriaceae; Salmonella.
OX	NCBI_Taxid=602;
RN	[1]
RP	SEQUENCE FROM N.A.
RA	MEDLINE=98053981; PubMed=9393832;
RA	Sukhopolvi S.S., Lorentz R.G., Gordon J.I., Bian Z., Pfeiffer J.D.,
RA	Normark S.J., Rhen M.;
RT	"Expression of thin, aggregative fimbriae promotes interaction of
RT	Salmonella typhimurium SR-11 with mouse small intestinal epithelial
RT	cells.";
RL	Infect. Immun. 65:5320-5325(1997).
DR	EMBL; AJ000514; CAA04151.1; -
FT	NON_TER 152 152
SO	SEQUENCE 152 AA; 15401 MW; 9DA7DADC2364B006 CRC64;
Query Match	87.4%; Score 682; DB 2; Length 152;
Best Local Similarity	89.4%; Pred. No. 1.7e-46;
Matches 135; Conservative	4; Mismatches 12; Indels 0; Gaps 0;
QY	1 MLLKLVAAFAAIVSGSALAGVPPWGGGCGNNGGSGSPDYDOLYTRVTHENHALQ 60
DB	1 MLLKLVAAFAAIVSGSAVAVPWPWGGGCGNNGGSGSPDSTISIVYGSNNALALQ 60
QY	61 SPARKSETTITSGYNGADVQAGADNRTIELTONGFRNNATIDQNNAKNSDITVGOYG 120
DB	61 SPARKSETTITSGYNGADVQAGADNRTIELTONGFRNNATIDQNNAKNSDITVGOYG 120
QY	121 NNAALVNOTASDSSVMVROVFGNNATANQY 151

```

Db      121 NNAALVNOTASDSSVWVRQVGFGNNAIPANQY 151
      |||
RESULT 2
QX243  PRELIMINARY; PRT; 150 AA.
AC  Q7X243;
DT  01-OCT-2003 (TREMBlrel. 25, Created)
DT  01-OCT-2003 (TREMBlrel. 25, Last sequence update)
DT  01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE  Curliin-csgA protein.
GN  CSGA.
OS  Citrobacter sp. Fec2.
OC  Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC  Enterobacteriaceae; Citrobacter.
OX  NCBI_TaxId=211763;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=Fec2;
RA  Zogaj X., Bokranz W., Nimtz M., Romling U.;
RT  "Production of Cellulose and Curli Fimbriae by Members of the Family
    Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL  Infect. Immun. 72:4151-4158 (2003).
DR  EMBL; AJ515700; CAD56672.1; -
SQ  SEQUENCE 150 AA; 15016 MW; 1D71418D6973DC6 CRC64;

Query Match      76.1%; Score 593.5; DB 2; Length 150;
Best Local Similarity 76.1%; Pred. No. 1.7e-39;
Matches 118; Conservative 14; Mismatches 18; Indels 1; Gaps 1;

QY  1 MLLKVAAPFAIIVSGSALAGVTPWVGCGNHNGCGSSGPDYDQLVTRVVTHEMAHALQ 60
Db  1 MLLQVAPFAIIVSGSALAGVTPWVGCGCGG-GGGGSSSGESTLSIYGVNNAIALQ 59
QY  61 SDARKSETTITQSGYNGADVGGADNSTIELTONGFRNNATTIDWNAKNSDITVGGYGG 120
Db  60 SDARKSDTTTHONGFNGADVGGQSDNSTIDLTONGFRNNATTIDWNAKNSDITVGGYGG 119
QY  121 NNAALVNOTASDSSVWVRQVGFGNNAIPANQY 151
Db  120 NNAALVNOTASDSSVWVRQVGFGNNAIPANQY 150

RESULT 3
QX240  PRELIMINARY; PRT; 149 AA.
AC  Q7X240;
DT  01-OCT-2003 (TREMBlrel. 25, Created)
DT  01-OCT-2003 (TREMBlrel. 25, Last sequence update)
DT  01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE  Curliin-csgA protein.
GN  CSGA.
OS  Citrobacter freundii.
OC  Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC  Enterobacteriaceae; Citrobacter.
OX  NCBI_TaxId=346;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=Fec4;
RA  Zogaj X., Bokranz W., Nimtz M., Romling U.;
RT  "Production of Cellulose and Curli Fimbriae by Members of the Family
    Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL  Infect. Immun. 72:4151-4158 (2003).
DR  EMBL; AJ515701; CAD56675.1; -
SQ  SEQUENCE 149 AA; 15260 MW; 946D52017F648FD CRC64;

Query Match      70.5%; Score 550; DB 2; Length 149;
Best Local Similarity 72.2%; Pred. No. 4.5e-36;
Matches 109; Conservative 20; Mismatches 20; Indels 2; Gaps 1;

QY  1 MLLKVAAPFAIIVSGSALAGVTPWVGCGNHNGCGSSGPDYDQLVTRVVTHEMAHALQ 60
      |||

```

```

Db      1 MLLKVAAPFAIIVSGSALAGVTPWVGCGNHNGCGSSGPDYDQLVTRVVTHEMAHALQ 58
QY  61 SDARKSETTITQSGYNGADVGGADNSTIELTONGFRNNATTIDWNAKNSDITVGGYGG 120
Db  59 SDARKSDTTTHONGFNGADVGGADNSTIELTONGFRNNATTIDWNAKNSDITVGGYGG 118
QY  121 NNAALVNOTASDSSVWVRQVGFGNNAIPANQY 151
Db  119 NNAALVNOTASDSSVWVRQVGFGNNAIPANQY 149

RESULT 4
QX243  PRELIMINARY; PRT; 152 AA.
AC  Q8CW63;
DT  01-MAR-2003 (TREMBlrel. 23, Created)
DT  01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT  01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE  Major curliin subunit precursor.
GN  CSGA OR C1306.
OS  Escherichia coli O6.
OC  Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC  Enterobacteriaceae; Escherichia.
OX  NCBI_TaxId=217992;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=O6:HT / ATCC 700928;
RX  MEDLINE=22386234; PubMed=12471157;
RA  Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA  Raske D., Buckles E.L., Lion S.-R., Boutin A., Hackett J., Stroud D.,
RA  Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA  Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT  "Extensive mosaic structure revealed by the complete genome sequence
    of uropathogenic Escherichia coli.";
RL  Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).
DR  EMBL; AE016759; AAN79779.1; -
KW  Complete proteome.
SQ  SEQUENCE 152 AA; 15064 MW; 3BA57F34C1240E83 CRC64;

Query Match      63.9%; Score 498.5; DB 16; Length 152;
Best Local Similarity 67.1%; Pred. No. 5.4e-32;
Matches 102; Conservative 20; Mismatches 29; Indels 1; Gaps 1;

QY  1 MLLKVAAPFAIIVSGSALAGVTPWVGCGNHNGCGSSGPDYDQLVTRVVTHEMAHALQ 59
Db  1 MLLKVAAPFAIIVSGSALAGVTPWVGCGNHNGCGSSGPDYDQLVTRVVTHEMAHALQ 60
QY  60 QSDARKSETTITQSGYNGADVGGADNSTIELTONGFRNNATTIDWNAKNSDITVGGYGG 119
Db  61 QADARNSDLTITQHGNGADVGGQSDNSTIDLTORGFGNSATIDWNAKNSDITVGGYGG 120
QY  120 GNNALVNOTASDSSVWVRQVGFGNNAIPANQY 151
Db  121 GNGAIVDQTNSSSVNVTQVGFGNNAIPANQY 152

RESULT 5
QX237  PRELIMINARY; PRT; 150 AA.
AC  Q7X237;
DT  01-OCT-2003 (TREMBlrel. 25, Created)
DT  01-OCT-2003 (TREMBlrel. 25, Last sequence update)
DT  01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE  Curliin-csgA protein.
GN  CSGA.
OS  Enterobacter sakazakii.
OC  Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC  Enterobacteriaceae; Enterobacter.
OX  NCBI_TaxId=28141;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=Fec39;
RA  Zogaj X., Bokranz W., Nimtz M., Romling U.;

```

RT "Production of Cellulose and Curli Fimbriae by Members of the Family
 Enterobacteriaceae Isolated from the Human Gastrointestinal Tract."
 Infect. Immun. 72:4151-4158(2003).
 RL EMBL; AJ515702; CAD56678.1; -
 SQ SEQUENCE 150 AA; 15112 MW; 5DB8BD2872DF15F3 CRC64;

Query Match 54.9%; Score 428.5; DB 2; Length 150;
 Best Local Similarity 55.1%; Pred. No. 1.8e-26;
 Matches 92; Conservative 21; Mismatches 21; Indels 33; Gaps 3;

QY 1 MLKLKVAAPFAIVSGSALAGVVPQ--WGCGGNH-----NGGNSGSPDYD 44
 DB 1 MRFIRKALALAIIVSSAMAGMINCGKGGHGGGCGPNSLTINIQNGGNS----- 54
 QY 45 QVTRVVTHEMALQSDARKSETTITGSGYGNGADVGQADNSTIELTONGFRNATID 104
 DB 55 -----ALALGTDAKNSVLTNISQCGGNGADVGQSDDSINTLQNGGNSATID 103

QY 105 QNNAKNSDITVGQYGGNNALVNGTASDSVWVRQVGFNNATANQY 151
 DB 104 QNNSKDSVWVNSQYGGNGALVDQTASNSTVNTVQIGFGRHATANQY 150

RESULT 6

Q54069 PRELIMINARY; PRT; 76 AA.

AC Q54069; 01-NOV-1996 (TRENBLREL. 01, Created)
 DT 01-NOV-1996 (TRENBLREL. 01, Last sequence update)
 DT 01-DEC-2001 (TRENBLREL. 19, Last annotation update)
 DE SEF17 fimbria (Fragment).
 GN AGFA.
 OS Salmonella enteritidis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 RX NCBI_Taxid=592;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SE30;
 RA Cox J.M., Eglesos S., Woolcock J.B.;
 RT "Virulence of Salmonella enteritidis in chickens correlates with
 colony morphology and expression of SEF17 fimbriae."
 RL Submitted (Apr-1996) to the EMBL/Genbank/DBD databases.
 RL EMBL; U53207; AAA8671.1; -
 FT NON TER 1 1
 FT NON TER 76 76
 SQ SEQUENCE 76 AA; 7704 MW; 2FDS411241A7BCB1 CRC64;

Query Match 38.8%; Score 303; DB 2; Length 76;
 Best Local Similarity 80.3%; Pred. No. 6.6e-17;
 Matches 61; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

QY 30 GNNHGGSSGPDYDQVTRVVTHEMALQSDARKSETTITGSGYGNGADVGQADNST 69
 DB 1 GNNHGGSSGPDYDQVTRVVTHEMALQSDARKSETTITGSGYGNGADVGQADNST 60

QY 90 IELTONGFRNATIDQ 105
 DB 61 IELTONGFRNATIDQ 76

RESULT 7

Q9S3J5 PRELIMINARY; PRT; 29 AA.

AC Q9S3J5; 01-MAY-2000 (TRENBLREL. 13, Created)
 DT 01-MAY-2000 (TRENBLREL. 13, Last sequence update)
 DT 01-MAR-2001 (TRENBLREL. 16, Last annotation update)
 DE Curliin subunit monomer (Fragment).
 GN CSGA.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.

OX NCBI_Taxid=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TRANSPOSON-insertion sequence IS1;
 RX MEDLINE=99314153; PubMed=10386375;
 RA La Ragione R.M., Collichan R.J., Woodward M.J.;
 RT "Non-curliation of Escherichia coli O78:H80 isolates associated with
 ISI insertion on in csgB and reduced persistence in poultry infection."
 RL FEMS Microbiol. Lett. 175:247-253(1999).
 RL EMBL; AJ131756; CAB45380.1; -
 FT NON TER 29 29
 SQ SEQUENCE 29 AA; 2789 MW; E29DPC07ABBB243 CRC64;

Query Match 15.6%; Score 122; DB 2; Length 29;
 Best Local Similarity 89.7%; Pred. No. 0.0042;
 Matches 26; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MLKLKVAAPFAIVSGSALAGVVPQYGGG 29
 DB 1 MLKLKVAAPFAIVSGSALAGVVPQYGGG 29

RESULT 8

Q94821 PRELIMINARY; PRT; 1748 AA.

AC Q94821; P92146; P92145; P92144; P92142; P92141; Q94820;
 DT 01-FEB-1997 (TRENBLREL. 02, Created)
 DT 01-FEB-1997 (TRENBLREL. 02, Last sequence update)
 DT 01-OCT-2003 (TRENBLREL. 25, Last annotation update)
 DE CNUB protein.
 GN CNUB.
 OS Tetrahymena thermophila.
 OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
 OC Tetrahymena; Tetrahymena.
 RX NCBI_Taxid=5911;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=88189811; PubMed=3357771;
 RA Martindale D.W., Taylor F.M.;
 RT "Multiple introns in a conjugation-specific gene from Tetrahymena
 thermophila."
 RL Nucleic Acids Res. 16:2189-2201(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=94051569; PubMed=823798;
 RA Taylor F.M., Martindale D.W.;
 RT "Retroviral-type zinc fingers and glycine-rich repeats in a protein
 encoded by cnuB, a Tetrahymena gene active during meiosis."
 RL Nucleic Acids Res. 21:4610-4614(1993).
 RL EMBL; X06462; CAB37323.1; -
 DR EMBL; L03710; AAC37171.1; -
 DR PIR; S42136; S42136.
 DR HSP; P05886; 1AAP.
 DR GO; GO:0003676; R-nucleic acid binding; IEA.
 DR InterPro; IPR001878; Znf_CCHC.
 DR Pfam; PF00098; zF_CCHC; 7.
 DR PRINTS; PR00939; C2HCZNFINGER.
 DR SMART; SM00343; Znf_C2HC; 7.
 DR PROSITE; PS50158; zF_CCHC; 7.
 FT CONFLICT 251 251 M -> I (IN REF. 1).
 FT CONFLICT 256 256 I -> N (IN REF. 1).
 SQ SEQUENCE 1748 AA; 199624 MW; 0B03F210104008A3 CRC64;

Query Match 14.6%; Score 114; DB 5; Length 1748;
 Best Local Similarity 30.5%; Pred. No. 2.3; 27; Indels 40; Gaps 8;
 Matches 39; Conservative 22; Mismatches 2; Indels 40; Gaps 8;

QY 25 QWGGGNNHNGG--GNSGPDYDQVTRVVTHEMALQSDARKSETTITGSGYGNGADV 81
 DB 1640 QWGGGNNHNGG--GNSGPDYDQVTRVVTHEMALQSDARKSETTITGSGYGNGADV 1680

QY 82 GGGADNSTIELTONGFRNATIDQNNAKNSDITVGQYGGNNALVNGTASDSVWVRQV 141

Db 1681 -SGSGN-----TGGGMSN---DNOOQONNTGGGSSNS---NOTNESS----- 1722

Qy 142 FGNMNTAN 149

Db 1723 WGSNNQAS 1730

RESULT 9

Q8E1H4 PRELIMINARY; PRT; 502 AA.

AC Q8E1H4; 01-MAR-2003 (TREMBlrel. 23, Created)

DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)

DE 01-MAR-2003 (TREMBlrel. 23, Last annotation update)

DE Conserved hypothetical protein.

GN S00865.

OS Shewanella oneidensis.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;

OC Alteromonadaceae; Shewanella.

OX NCBI_TaxID=70863;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=MR-1;

RX MEDLINE=22297686; PubMed=12368813;

RA Heidelberg J.F., Paulsen I.T., Neilson K.E., Gaidos E.J., Neilson W.C.,

RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,

RA Meyer T., Tsapin A., Scott J., Benham M., Brinkac L., Daugherty S.,

RA Deboy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,

RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,

RA Vamathevan J., Weidman J., Imprimis M., Lee K., Berry K., Lee C.,

RA Mueller T.V., Smith H.O., Venter J.C., Neilson K.H., Fraser C.M.,

RT "Genome sequence of the dissimilatory metal ion-reducing bacterium

RT Shewanella oneidensis."

RL Nat. Biotechnol. 20:1118-1123(2002).

DR EMBL; A015532; AAN53941.1; .

DR TIGR; S00865; .

KW Hypothetical protein; Complete proteome.

SO SEQUENCE 502 AA; 52441 MW; D08CA23D6C4B62D CRC64;

Query Match 14.5%; Score 113; DB 16; Length 502;
Best Local Similarity 26.0%; Pred. No. 0.64;
Matches 40; Conservative 20; Mismatches 58; Indels 36; Gaps 6;

Qy 29 GGNNG-----GGN-----SSGPDYQVLR-----VTHMAHAL 59

Db 231 GDNHTGVYALAGSNDISHEQGSNTAYLSMTTGDNTVDITQDGSNTVQDSLADI 290

Qy 60 QSDARKSETTITSGYNGADVGQADNSTIELTONGFRNNATIDQWNAKNSDITVQYV 119

Db 291 QGD--DNDITIKQKGSNGAFQVWGDSDNDVLDKQGDANFAFGAVGTDN-DPDLSSKG 347

Qy 120 GNNALVNOTASDSVWVRQVFGN-----NATAN 149

Db 348 DNNELVAFATGEDNSIETISQGDANFAYVATGDN 381

RESULT 10

Q7X244 PRELIMINARY; PRT; 151 AA.

AC Q7X244; 01-OCT-2003 (TREMBlrel. 25, Created)

DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)

DE 01-OCT-2003 (TREMBlrel. 25, Last annotation update)

DE Nucleation component of curlin monomers.

GN C5GB.

OS Citrobacter sp. Fec2.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Citrobacter.

OX NCBI_TaxID=213763;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Fec2;

RA Zogaj X., Bokranz W., Nintz M., Romling U.;
RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RT Enterobacteriaceae Isolated from the Human Gastrointestinal Tract."
RL Infect. Immun. 72:4151-4158(2003).
DR EMBL; AJ515700; CAD56571.1; .
SO SEQUENCE 151 AA; 16158 MW; BD00AF57E1400704 CRC64;

Query Match 14.2%; Score 110.5; DB 2; Length 151;
Best Local Similarity 25.0%; Pred. No. 0.24;
Matches 29; Conservative 26; Mismatches 46; Indels 15; Gaps 3;

Qy 47 VTRVTHMAHALQSPARKSETTITSGYNGADVGQADNSTIELTONGFR----- 98

Db 18 IASATYDLAH---SEYNFANVELSKSPNOAIIQGVETNNKAKRQDGSKLLSVSOE 74

Qy 99 ---NNATIDQWNAKNSDITVQYNGNNAALVNOTASDSVWVRQVFGNNTANQY 151

Db 75 GGSNRAKVQSGAYNF-AVIAQSGHNDASISQSNVGNMTIQQSGNKAITQY 129

RESULT 11

Q89J13 PRELIMINARY; PRT; 171 AA.

AC Q89J13; 01-JUN-2003 (TREMBlrel. 24, Created)

DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)

DE 01-JUN-2003 (TREMBlrel. 24, Last annotation update)

DE CsgA protein.

GN CSGA OR BL15300.

OS Bradyrhizobium japonicum.

OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;

OC Bradyrhizobiaceae; Bradyrhizobium.

OX NCBI_TaxID=375;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=USDA 110;

RX MEDLINE=22484998; PubMed=12597275;

RA Kaneo T., Nakamura Y., Sato S., Minamisawa K., Uchiyama T.,

RA Sasamoto S., Watanabe A., Ideawa K., Itiguchi M., Kawashima K.,

RA Kohara M., Matsunoto M., Shimo S., Tsuruko H., Wada T., Yamada M.,

RA Tabata S.;

RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium

RT Bradyrhizobium japonicum USDA110."

RL DNA Res. 9:189-197(2002).

DR EMBL; AP005954; BAC50565.1; .

KW Complete proteome.

SO SEQUENCE 171 AA; 17448 MW; 995DB08C01498381 CRC64;

Query Match 13.8%; Score 108; DB 16; Length 171;
Best Local Similarity 36.6%; Pred. No. 0.44;
Matches 34; Conservative 10; Mismatches 39; Indels 10; Gaps 3;

Qy 58 ALQSDARKSETTITSGYNGADVGQADNSTIELTONGFRNNATIDQWNAKNSDITV 115

Db 52 ALSSAQAANSTVQVGLVNGSSVTVQGLTNDSSSTQIGLNGASTMQGTSSPS----- 107

Qy 116 GQYGGNNAALVNOTASDSVWVRQVFGNATA 148

Db 108 ---LNNVSTVQAGVQNSATTGQVAFGNNSA 136

RESULT 12

Q7UC21 PRELIMINARY; PRT; 151 AA.

AC Q7UC21; 01-OCT-2003 (TREMBlrel. 25, Created)

DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)

DE 01-OCT-2003 (TREMBlrel. 25, Last annotation update)

DE Minor curlin subunit.

GN C5GB OR S1108.

OS Shigella flexneri.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Shigella.

OK NCBI_TaxId=623;
RN (1)
RC SEQUENCE FROM N.A.
RX STRAIN=2457T / ATCC 700930 / Serotype 2a;
MEDLINE=22590274; PubMed=12704152;
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
Schwartz D.C., Blattner F.R.;
RT "Complete genome sequence and comparative genomics of Shigella
flexneri serotype 2a strain 2457T.";
RL Infect. Immun. 71:2775-2786(2003).
DR EMBL: AB016981; AAF16542.1; -
SQ SEQUENCE 151 AA; 15868 MW; 5D5D266B964014A0 CRC64;
Query Match 13.8%; Score 107.5; DB 16; Length 151;
Best Local Similarity 35.4%; Pred. No. 0.42; Mismatches 43; Indels 1; Gaps 1;
Matches 29; Conservative 9;
QY 70 ITQSGYGNGADVQAGDNTIELTONGFRNNATIDQMAKNSDITVGQYGGNNALVNOT 129
DB 49 IQAGTNNNSAQLKQGSKLAVVAQGSNSRAKIDQTGPDYNL-AVIDQGSANDASISOG 107
QY 130 ASDSSVMVRQVGFNNATANOY 151
DB 108 AYGNTAMIIQKSGSKNKANITQY 129
RESULT 13
Q86C64 PRELIMINARY; PRT; 160 AA.
AC Q86C64;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Minor curlin subunit precursor.
GN CSGO OR C1305.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OK NCBI_TaxId=217992;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
Raeko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
DR EMBL: AB016759; AAN79778.1; -
KW Complete proteome.
SQ SEQUENCE 160 AA; 16963 MW; 49F68448D979B986 CRC64;
Query Match 13.8%; Score 107.5; DB 16; Length 160;
Best Local Similarity 35.4%; Pred. No. 0.45; Mismatches 43; Indels 1; Gaps 1;
Matches 29; Conservative 9;
QY 70 ITQSGYGNGADVQAGDNTIELTONGFRNNATIDQMAKNSDITVGQYGGNNALVNOT 129
DB 58 IQAGTNNNSAQLKQGSKLAVVAQGSNSRAKIDQTGPDYNL-AVIDQGSANDASISOG 116
QY 130 ASDSSVMVRQVGFNNATANOY 151
DB 117 AYGNTAMIIQKSGSKNKANITQY 138
RESULT 14
Q83R07 PRELIMINARY; PRT; 160 AA.
AC Q83R07;
DB 934 YGNVTVTGSFGATTAL-----GNGNDTIDASG--YGNVITLNGNDIVHPGASGQTGAG 987

DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Minor curlin subunit precursor, similar to CsgA.
GN CSGO OR SFI035.
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OK NCBI_TaxId=623;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=301 / Serotype 2a;
RX MEDLINE=22272406; PubMed=12384590;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
Yu J.;
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
through comparison with genomes of Escherichia coli K12 and O157.";
RL Nucleic Acids Res. 30:4432-4441(2002).
DR EMBL: AB015131; AAN42658.1; -
KW Complete proteome.
SQ SEQUENCE 160 AA; 16919 MW; 50269F5268D2A32F CRC64;
Query Match 13.8%; Score 107.5; DB 16; Length 160;
Best Local Similarity 35.4%; Pred. No. 0.45; Mismatches 43; Indels 1; Gaps 1;
Matches 29; Conservative 9;
QY 70 ITQSGYGNGADVQAGDNTIELTONGFRNNATIDQMAKNSDITVGQYGGNNALVNOT 129
DB 58 IQAGTNNNSAQLKQGSKLAVVAQGSNSRAKIDQTGPDYNL-AVIDQGSANDASISOG 116
QY 130 ASDSSVMVRQVGFNNATANOY 151
DB 117 AYGNTAMIIQKSGSKNKANITQY 138
RESULT 15
Q89CK5 PRELIMINARY; PRT; 1209 AA.
AC Q89CK5;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE BL17792 protein.
GN BL17792.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OK NCBI_TaxId=375;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RX MEDLINE=22484998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiyama T.,
RA Sasamoto S., Watanabe A., Idegawa K., Iriyuchi M., Kawashima K.,
RA Kohara M., Matsunoto M., Shimpo S., Teunroka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
Bradyrhizobium japonicum USDA110.";
RL DNA Res. 9:189-197(2002).
DR EMBL: AP005963; BAC53057.1; -
KW Complete proteome.
SQ SEQUENCE 1209 AA; 118292 MW; 50F97581D524EB71 CRC64;
Query Match 13.7%; Score 106.5; DB 16; Length 1209;
Best Local Similarity 25.0%; Pred. No. 5.9; Mismatches 47; Conservative 20; Mismatches 66; Indels 55; Gaps 8;
Matches 47;
QY 9 FAIVYSGALAGVPRQWCGGNGHNGGSGPDYQVTVVTHMAH-----ALQSDR 64
DB 934 YGNVTVTGSFGATTAL-----GNGNDTIDASG--YGNVITLNGNDIVHPGASGQTGAG 987

Qy	65	KSETTITOSGYNGADVGGAD-----NSTIELTONGFRNATIDQNN-	107
Db	988	NGNDLVTLISGYNTVMLENGNDVAGRDGANSVTLGDGNTVNL--GCMGQITVSGGTN	1045
Qy	108	---AKNSDITVGOYG-----GNNALVNO-----TASDSVWVROVG	141
Db	1046	AIAGSGSDNVVAGAGHPTIMLGAAHVVLNGSQANTNOIGQDVTVNGSGSDQNFVG	1105
Qy	142	FGNNATAN	149
Db	1106	FGNOAIIN	1113

Search completed: March 11, 2004, 18:40:33
 Job time : 49.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: March 11, 2004, 18:13:53 ; Search time 45.9 Seconds
(without alignments)
929.514 Million cell updates/sec

Title: US-09-543-407-20

Perfect score: 774
Sequence: 1 MLLKLVAAFAALVSGSALA.....DSSVWVROYFGNNATANY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*

- 1: geneseqp19808:*
- 2: geneseqp19908:*
- 3: geneseqp20008:*
- 4: geneseqp20018:*
- 5: geneseqp20028:*
- 6: geneseqp20038:*
- 7: geneseqp20038:*
- 8: geneseqp20048:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	774	100.0	151	3	AAB36350 Agfa::PT3
2	709	91.6	151	3	AAB36353 Agfa::PT3
3	689	89.0	151	2	AAB74625 Agfa sequ
4	689	89.0	151	3	AAB36341
5	684	88.4	151	2	AAB23570 Salmonell
6	662	85.5	151	3	AAB36354 Agfa::PT3
7	622	80.4	151	3	AAB36349 Agfa::PT3
8	619	80.0	151	3	AAB36351 Agfa::PT3
9	611	78.9	151	3	AAB36346 Agfa::PT3
10	609	78.7	151	3	AAB36347 Agfa::PT3
11	606	78.3	151	3	AAB36352 Agfa::PT3
12	597	77.1	151	3	AAB36355 Agfa::PT3
13	574	74.2	151	3	AAB36348 Agfa::PT3
14	528	68.2	151	7	AAB36343 Escherich
15	523	67.6	151	7	ABR82651 E. coli C
16	504	65.1	120	2	AAB62761 Agfa sequ
17	504	65.1	120	2	AAB23568 Salmonell
18	450	58.1	142	2	AAB2664 Fibronect
19	378	48.8	122	2	AAB2663 FNB curli
20	153	19.8	45	3	AAB36316 Salmonell
21	132	17.1	22	3	AAB36318 Salmonell
22	123	15.9	23	3	AAB36321 Salmonell
23	123	15.9	23	3	AAB36326 Salmonell
24	123	15.9	23	3	AAB36338 Salmonell
25	113	14.6	24	7	ABR82644 E. coli C

26	111	14.3	22	3	AAB36322
27	111	14.3	22	3	AAB36327
28	111	14.3	22	3	AAB36337
29	109	14.1	23	3	AAB36340
30	109	14.1	23	3	AAB36324
31	109	14.1	23	3	AAB36319
32	102	13.2	26	7	ABR82649
33	96.5	12.5	151	3	AAB36344
34	96	12.4	19	3	AAB36323
35	96	12.4	19	3	AAB36336
36	95	12.4	19	3	AAB36328
37	95	12.3	24	7	ABR82647
38	92	11.9	23	3	AAB36331
39	91	11.8	186	6	ABU21488
40	91	11.8	502	2	AAW32312
41	90.5	11.7	677	4	ABG04318
42	90	11.6	24	7	ABR82642
43	89.5	11.6	423	4	ABG07164
44	89.5	11.6	447	3	ABG29728
45	89.5	11.6	468	3	ABG29727

ALIGNMENTS

RESULT 1	AAB36350	standard; protein; 151 AA.
ID	AAB36350	
XX		
AC	AAB36350;	
XX		
DT	26-FEB-2001 (first entry)	
XX		
DE	Agfa::PT3#5 amino acid sequence SEQ ID NO:20.	
XX		
KW	Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;	
XX	vaccine; immune response; immunogen.	
OS	Salmonella enteritidis.	
OS	Escherichia coli.	
XX	Synthetic.	
PN	WO200060102-A2.	
XX		
PD	12-OCT-2000.	
XX		
PF	05-APR-2000; 2000MO-CA000356.	
XX		
PR	05-APR-1999; 99US-0127888P.	
XX		
PA	(UUYI-) UNIV VICTORIA.	
XX		
PI	White AP, Doran JL, Collison SK, Kay WW;	
XX		
DR	WPI; 2000-672631/65.	
XX		
NR	N-PSDB; AAC64626.	
PT	Recombinant agfa gene having a segment replaced by a foreign DNA sequence	
XX	which encodes foreign epitope or antigen, expresses recombinant Agfa	
PT	protein useful for eliciting immune response in animal.	
XX		
PS	Disclosure; Page 137; 139pp; English.	
XX		
CC	The present invention describes a recombinant agfa gene (1) where a	Aab36322 Salmonell
CC	segment of the gene has been replaced by a segment of a foreign DNA	Aab36337 Salmonell
CC	sequence which encodes a foreign epitope or antigen. Also described are:	Aab36337 Salmonell
CC	(1) use of thin aggregative fimbriae (Sfp17/TAF) nucleation depended	Aab36340 Salmonell
CC	assembly system of strains of Salmonella, Escherichia coli and	Aab36324 Salmonell
CC	Enterobacteriaceae for the production of fimbriae comprising recombinant	Aab36319 Salmonell
CC	Agfa, Cga and Agfa-homologue fimbria subunits, respectively; (2)	ABR82649 E. coli V
CC	directing recombination of a recombinant gene into the chromosome of the	Aab36344 Escherich
CC	homologous species; (3) directing recombination of a recombinant gene	Aab36336 Salmonell
CC	back into the chromosome of the homologous species, replacing the native	Aab36328 Salmonell
CC		ABR82647 E. coli C
CC		Aab36331 Escherich
CC		ABU21488 Protein e
CC		AAW32312 Leishman
CC		ABG04318 Novel hum
CC		ABR82642 E. coli N
CC		ABG07164 Novel hum
CC		ABG29728 Arabidops
CC		ABG29727 Arabidops

CC	copy of that gene; and (4) eliciting an immune response in an animal,
CC	comprising separating an amino acid polymer comprising a recombinant Agfa
CC	protein containing a replacement segment or segments of foreign amino
CC	acid sequence or sequences grown on a Salmonella, E. coli or
CC	Enterobacteriaceae host cell, from the host cell and introducing the
CC	polymer into the animal in conjunction with a carrier or diluent. (I) is
CC	useful for the expression of recombinant Agfa protein which is useful for
CC	eliciting an immune response in an animal. In a fibrillar presentation
CC	system the heterologous antigens are presented in high numbers (up to
CC	500,000 copies/cell), the hybrid fibrin protein possesses both the
CC	immunogenicity and adhesion properties relevant for an efficient live
CC	vaccine, the carrier fibrillar subunit proteins are usually strong
CC	immunogens, which may be important for directing an immune response
CC	against the inserted epitope, and hybrid fibrillae are easy and
CC	inconspicuous to purify in large amount. The present sequence is given in
CC	the exemplification of the present invention
XX	
SQ	Sequence 151 AA;
Query Match	100.0%; Score 774; DB 3; Length 151;
Best Local Similarity	100.0%; Pred. No. 8,8e-71;
Matches 151; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
OY	1 MKLKVAAPAIVVSGSALAGVVPOMGGGHNHNGSSSPDSTLSIYGSAANAALAQ 60
Dd	1 MKLLKVAFAAIVVSGSALAGVVPOMGGGHNHNGSSSPDSTLSIYGSAANAALAQ 60
OY	SDARKYDGLTVRVVTHEMAHGOGADNSTITELTONGFPPNNATITDDMNKNSDITYGYGG 120
Dd	SDARKYDGLTVRVVTHEMAHGOGADNSTITELTONGFPPNNATITDDMNKNSDITYGYGG 120
OY	121 NNAALVNQTASDSVWVRQVGFGNNAATNOY 151
Dd	121 NNAALVNQTASDSVWVRQVGFGNNAATNOY 151
RESULT 2	
AAB36353	
ID AAB36353 standard; protein; 151 AA.	
AC AAB36353;	
XX 26-FEB-2001 (first entry)	
DE Agfa::PT3#8 amino acid sequence SEQ ID NO:26.	
KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;	
XV vaccine; immune response; immunogen.	
OS Salmonella enteritidis.	
OS Escherichia coli.	
OS Synthetic.	
PX WO20060102-A2.	
PD 12-OCT-2000.	
XX 05-APR-2000; 2000WO-CA000356.	
XX 05-APR-1999; 99US-0127888P.	
PA (UYVI-) UNIV VICTORIA.	
PI White AP, Doran JL, Collison SK, Kay WW,	
DR WPI; 2000-672631/65.	
XX N-PADB; AAC64629.	
PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence	
PT which encodes foreign epitope or antigen, expresses recombinant Agfa	
XX protein useful for eliciting immune response in animal.	
XX Disclosure; Page 138; 139pp; English.	

XX		The present invention describes a recombinant agfa gene (I) where a
CC		segment of the gene has been replaced by a segment of a foreign DNA
CC		sequence which encodes a foreign epitope or antigen. Also described are:
CC	(1)	use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC		assembly system of strains of Salmomella, Escherichia coli and
CC		Enterobacteriaceae for the production of fimbriae comprising recombinant
CC	Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)	
CC		directing recombination of a recombinant gene into the chromosome of the
CC		homologous species; (3) directing recombination of a recombinant gene
CC		back into the chromosome of the homologous species, replacing the native
CC		copy of that gene; and (4) eliciting an immune response in an animal,
CC		comprising separating an amino acid polymer comprising a recombinant Agfa
CC		protein containing a replacement segment or segments of foreign amino
CC		acid sequence or sequences grown on a Salmomella, E. coli or
CC		Enterobacteriaceae host cell, from the host cell and introducing the
CC		polymer into the animal in conjunction with a carrier or diluent. (I) is
CC		useful for the expression of recombinant Agfa protein which is useful for
CC		eliciting an immune response in an animal. In a fimbrial presentation
CC		system the heterologous antigens are presented in high numbers (up to
CC		500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC		immunogenicity and adhesion properties relevant for an efficient live
CC		vaccination, the carrier fimbrial subunit proteins are usually strong
CC		immogens, which may be important for directing an immune response
CC		against the inserted epitope, and hybrid fimbriae are easy and
CC		inexpensive to purify in large amount. The present sequence is given in
CC		the exemplification of the present invention
XX	SQ	Sequence 151 AA;
XX		
XX	Query Match	91.6%; Score 709; DB 3; Length 151;
XX	Best Local Similarity	89.9%; Pred. No. 3.7e-64;
XX	Matches 143; Conservative	0; Mismatches 0; Indels 16; Gaps 2;
OY	1	MKLKVAFAPIVWSGALAGVPOWGCGGHNGSGNSGGPDTLSITYOXSANAALALQ 60
DB	1	MKLAKNAFAPIIVSGSHLACGVFOWGGGNHNGSGNSGGPDTLSITIOYSANAL-- 57
OY	61	SDARKDOLVTRVVTHEMAA-----GGADNSTIELTONGFRNNATIDOWNAXKD 112
DB	58	----YDLVTRVVTHEMAHAHGYNMGADVGGADNSTIELTONGFRNNATIDOWNAXKD 112
OY	113	ITVGQYGGNNAALVNQTASDSVWRQYGFNNATANQY 151
DB	113	ITVGQYGGNNAALVNQTASDSVWRQYGFNNATANQY 151
RESULT 3		
AAR74625		
ID	AAR74625 standard; protein; 151 AA.	
AC	AAR74625;	
XX		
XX	25-MAR-2003 (revised)	
DT	26-JUN-1995 (first entry)	
XX		
DE	Agfa sequence.	
XX		
XX	Salmomella; Agfa; vaccine.	
OS	Salmomella.	
FN	WO9425598-A2.	
PD	10-NOV-1994.	
XX		
PF	26-APR-1994; 94WO-IB000207.	
PR	26-APR-1993; 93US-00054452.	
PA	(UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.	
XX	(KING/) KING J.	

PI Kay WM, Collinson SK, Clouthier SC, Doran JL;
 XX MPI; 1994-358275/44.
 DR N-PSDB; AAC67467.
 XX
 PT Eliciting an immune response to *Salmonella* - using attenuated *Salmonella*
 PT strains, vector constructs, or compsns. contg. fimbrial type proteins.
 XX
 PS Disclosure; Fig 7B; 95pp; English.
 CC The *Salmonella* Agfa protein and DNA are used in vaccine and generic
 CC immunization compositions, respectively, to elicit an immune response to
 CC *Salmonella* in animals (e.g. food producing animals) and humans. (Updated
 CC on 25-MAR-2003 to correct PW field.)
 CC
 SQ Sequence 151 AA;
 Query Match 89.0%; Score 689; DB 2; Length 151;
 Best Local Similarity 90.7%; Pred. No. 4.1e-62;
 Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;
 QY 1 MKLLKVAAPFAIVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYQYGSANALALQ 60
 DB 1 MKLLKVAAPFAIVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYQYGSANALALQ 60
 QY 61 SDARKYDQLVTRVTHEMAHAGGADNSTIELTONGFRNNATIDQNNKNSDITVGOYGG 120
 DB 61 SDARKSETTITQSGYNGADVQGADNSTIELTONGFRNNATIDQNNKNSDITVGOYGG 120
 QY 121 NNAALVNQTPASDSSVMVRQVGFNNATANY 151
 DB 121 NNAALVNQTPASDSSVMVRQVGFNNATANY 151
 RESULT 4
 ID AAB36341 standard; protein; 151 AA.
 XX
 AC AAB36341;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE *Salmonella* enteritidis Agfa amino acid sequence SEQ ID NO:5.
 XX
 KM *Salmonella*; agfa; chromosomal gene replacement; fimbria; epitope;
 KM vaccine; immune response; immunogen.
 XX
 OS *Salmonella* enteritidis.
 OS
 PN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UUYI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collinson SK, Kay WM;
 XX
 DR MPI; 2000-672631/65.
 DR N-PSDB; AAC64617.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 135; 139pp; English.
 CC The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:

CC (1) use of thin aggregative fimbriae (SEF1/7AF) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, Caga and Agfa-homologue fimbria subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;
 Query Match 89.0%; Score 689; DB 3; Length 151;
 Best Local Similarity 90.7%; Pred. No. 4.1e-62;
 Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;
 QY 1 MKLLKVAAPFAIVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYQYGSANALALQ 60
 DB 1 MKLLKVAAPFAIVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYQYGSANALALQ 60
 QY 61 SDARKYDQLVTRVTHEMAHAGGADNSTIELTONGFRNNATIDQNNKNSDITVGOYGG 120
 DB 61 SDARKSETTITQSGYNGADVQGADNSTIELTONGFRNNATIDQNNKNSDITVGOYGG 120
 QY 121 NNAALVNQTPASDSSVMVRQVGFNNATANY 151
 DB 121 NNAALVNQTPASDSSVMVRQVGFNNATANY 151
 RESULT 5
 ID AAM23570 standard; protein; 151 AA.
 XX
 AC AAM23570;
 XX
 DT 25-MAR-2003 (revised)
 DT 29-SEP-1997 (first entry)
 XX
 DE *Salmonella* enteritidis 27655-3b agfa.
 XX
 KM Enteropathogenic bacteria; enterobacteria; *S. enteritidis*; antibody.
 KM *Salmonella* enteritidis.
 OS
 OS
 PN US5635617-A.
 XX
 PD 03-JUN-1997.
 XX
 PF 26-APR-1994; 94US-00233788.
 XX
 PR 26-APR-1993; 93US-00054452.
 XX
 PA (UUYI-) UNIV VICTORIA INNOVATION & DEV CORP.
 XX
 PI Collinson SK, Kay WM, Doran JL;

```

XX  WPI, 1997-309886/28.
DR  N-PSDB; AAT74142.
XX
PT  isolated Salmonella gene agfa - used for diagnosis of Salmonella or
XX  enteropathogenic bacteria of the Enterobacteria family.
XX
PS  Example 2; Fig 7; 85pp; English.
XX
CC  The present sequence represents agfa encoded by the full agfa gene
CC  derived from Salmonella enteritidis 27655-3b. The nucleic acid can be
CC  used to provide diagnostic assays for Salmonella and/or enteropathogenic
CC  bacteria of the family Enterobacteria. It can also be used to provide
CC  proteins and antibodies which can be used for assays. The nucleic acid
CC  sequence can be used to provide probes or primers which can specifically
CC  hybridise to nucleic acid molecules from greater than 99% of Salmonella
CC  strains that are pathogenic to warm-blooded animals relative to nucleic
CC  acid molecules from virtually all other microbial organisms. (Updated on
CC  25-MAR-2003 to correct PF field.)
XX
SQ  Sequence 151 AA;

Query Match      88.4%; Score 684; DB 2; Length 151;
Best Local Similarity 90.1%; Pred. No. 1.3e-61;
Matches 136; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

QY  1 MKLKVAFPAFAIVSGSALAGVFPQWGGGNNHNGSGSPDSTLSIYQGSANAALALQ 60
DB  1 MKLKVAFPAFAIVSGSALAGVFPQWGGGNNHNGSGSPDSTLSIYQGSANAALALQ 60
QY  61 SDARKITDQLVTRVVTHEMAHAGGADNSTIELTQNGFRNNATTIDQNAKNSDITVGQYG 120
DB  61 SDARKSETTITQSGYGNCADVDGCGADNSTIELTQNGFRNNATTIDQNAKNSDITVGQYG 120
QY  121 NNAALVNOTASDSSVMRVQGFNNATANOY 151
DB  121 NNPALVNOTASDSSVMRVQGFNNATANOY 151

RESULT 6
AAB36354
ID  AAB36354 standard; protein; 151 AA.
XX
AC  AAB36354;
XX
DT  26-FEB-2001 (first entry)
XX
DE  Agfa::PT3#9 amino acid sequence SEQ ID NO:28.
XX
KW  Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
XX  vaccine; immune response; immunogen.
XX
OS  Salmonella enteritidis.
OS  Escherichia coli.
OS  Synthetic.
XX
PN  WO200060102-A2.
XX
PD  12-OCT-2000.
XX
PF  05-APR-2000; 2000WO-CA000356.
XX
PR  05-APR-1999; 99US-0127888P.
XX
PA  (UYVI-) UNIV VICTORIA.
XX
PI  White AP, Doran UL, Collison SK, Kay WW;
XX  WPI, 2000-672631/65.
DR  N-PSDB; AAC64630.
XX
PT  Recombinant agfa gene having a segment replaced by a foreign DNA sequence
XX  which encodes foreign epitope or antigen, expresses recombinant Agfa

```

```

PT  protein useful for eliciting immune response in animal.
XX
PS  Disclosure; Page 138, 139pp; English.
XX
CC  The present invention describes a recombinant agfa gene (1) where a
CC  segment of the gene has been replaced by a segment of a foreign DNA
CC  sequence which encodes a foreign epitope or antigen. Also described are:
CC  (1) use of thin aggregative fimbriae (SEF71/7AP) nucleation depended
CC  assembly system of strains of Salmonella, Escherichia coli and
CC  Enterobacteriaceae for the production of fimbriae comprising recombinant
CC  Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
CC  directing recombination of a recombinant gene into the chromosome of the
CC  homologous species; (3) directing recombination of a recombinant gene
CC  back into the chromosome of the homologous species, replacing the native
CC  copy of that gene; and (4) eliciting an immune response in an animal,
CC  comprising separating an amino acid polymer comprising a recombinant Agfa
CC  protein containing a replacement segment or segments of foreign amino
CC  acid sequence or sequences grown on a Salmonella, E. coli or
CC  Enterobacteriaceae host cell, from the host cell and introducing the
CC  polymer into the animal in conjunction with a carrier or diluent. (1) is
CC  useful for the expression of recombinant Agfa protein which is useful for
CC  eliciting an immune response in an animal. In a fimbrial presentation
CC  system the heterologous antigens are presented in high numbers (up to
CC  500,000 copies/cell), the hybrid fimbria protein possesses both the
CC  immunogenicity and adhesion properties relevant for an efficient live
CC  vaccine, the carrier fimbrial subunit proteins are usually strong
CC  immunogens, which may be important for directing an immune response
CC  against the inserted epitope, and hybrid fimbriae are easy and
CC  inexpensive to purify in large amount. The present sequence is given in
CC  the exemplification of the present invention
XX
SQ  Sequence 151 AA;

Query Match      85.5%; Score 662; DB 3; Length 151;
Best Local Similarity 81.9%; Pred. No. 2.3e-59;
Matches 136; Conservative 0; Mismatches 0; Indels 30; Gaps 2;

QY  1 MKLKVAFPAFAIVSGSALAGVFPQWGGGNNHNGSGSPDSTLSIYQGSANAALALQ 60
DB  1 MKLKVAFPAFAIVSGSALAGVFPQWGGGNNHNGSGSPDSTLSIYQGSANAALALQ 60
QY  61 SPARK-----YDQLVTRVVTHEMAHAGGADNSTIELTQNGFRNNATTIDQ 105
DB  61 SPARKSETTITQSGYGNCADVDQLVTRVVTHEMAH-----FRNNATTIDQ 105
QY  106 WNAKNSDITVGQYGNNALVNOTASDSSVMRVQGFNNATANOY 151
DB  106 WNAKNSDITVGQYGNNALVNOTASDSSVMRVQGFNNATANOY 151

RESULT 7
AAB36349
ID  AAB36349 standard; protein; 151 AA.
XX
AC  AAB36349;
XX
DT  26-FEB-2001 (first entry)
XX
DE  Agfa::PT3#4 amino acid sequence SEQ ID NO:18.
XX
KW  Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
XX  vaccine; immune response; immunogen.
XX
OS  Salmonella enteritidis.
OS  Escherichia coli.
OS  Synthetic.
XX
PN  WO200060102-A2.
XX
PD  12-OCT-2000.
XX
PF  05-APR-2000; 2000WO-CA000356.
XX

```

PR 05-APR-1999; 99US-0127888P.
XX (UYVI-) UNIV VICTORIA.
PA White AP, Doran JL, Collison SK, Kay WM;
PI WPI; 2000-672631/65.
XX N-PSDB; AAC64625.
DR
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 136; 139pp; English.
XX
XX The present invention describes a recombinant agfa gene (1) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/7AF) nucleation depended
CC assembly system of strains of *Salmonella*, *Escherichia coli* and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
XX Sequence 151 AA:
SQ
Query Match 80.4%; Score 622; DB 3; Length 151;
Best Local Similarity 73.6%; Pred. No. 2.8e-55;
Matches 128; Conservative 0; Mismatches 0; Indels 46; Gaps 2;
QY 1 MLLKLVAAFAIVSGSALAGVPPQWGGGNNHNGSGPSTLSIYOGSANAALALQ 60
DB 1 MLLKLVAAFAIVSGSALAGVPPQWGGGNNHNGSGSPD----- 42
QY 61 SDARKVDQVTRVYTHEMAH-----GGAGNSTLTETONGP 97
DB 43 -----YDQVTRVYTHEMAHQSDARKSETTITGSGYNGADVGGADNSTLTETONGP 97
QY 98 RNNATIDOWNNAKNSDITVGQYGGNNAALVYQTSDDSVMVROYGFGNNAATANY 151
DB 98 RNNATIDOWNNAKNSDITVGQYGGNNAALVYQTSDDSVMVROYGFGNNAATANY 151
RESULT 8
AAB36351
ID AAB36351 standard; protein; 151 AA.
XX
XX AAB36351;
AC
XX
XX 26-FEB-2001 (first entry)
XX
XX Agfa::PTJ#6 amino acid sequence SEQ ID NO:22.
XX
XX Salmonella; agfa; chromosomal gene replacement; fimbrial; epitope;
KW vaccine; immune response; immunogen.

XX
XX *Salmonella enteritidis*.
OS *Escherichia coli*.
OS Synthetic.
XX
XX WO200060102-A2.
XX
XX 12-OCT-2000.
XX
XX 05-APR-2000; 2000WO-CA000356.
XX
XX
XX 05-APR-1999; 99US-0127888P.
XX
XX (UYVI-) UNIV VICTORIA.
PA White AP, Doran JL, Collison SK, Kay WM;
PI WPI; 2000-672631/65.
XX N-PSDB; AAC64627.
DR
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 137; 139pp; English.
XX
XX The present invention describes a recombinant agfa gene (1) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/7AF) nucleation depended
CC assembly system of strains of *Salmonella*, *Escherichia coli* and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
XX Sequence 151 AA:
SQ
Query Match 80.0%; Score 619; DB 3; Length 151;
Best Local Similarity 74.6%; Pred. No. 5.6e-55;
Matches 129; Conservative 0; Mismatches 0; Indels 44; Gaps 2;
QY 1 MLLKLVAAFAIVSGSALAGVPPQWGGGNNHNGSGPSTLSIYOGSANAALALQ 60
DB 1 MLLKLVAAFAIVSGSALAGVPPQWGGGNNHNGSGPSTLSIYOGSANAALALQ 60
QY 61 SDARK-----YDQVTRVYTHEMAHQSDARKSETTITGSGYNGADVGGADNSTLTETONGP 98
DB 61 SDARKSETTITGSGYNGADVGGADNDQVTRVYTHEMAH----- 103
QY 99 RNNATIDOWNNAKNSDITVGQYGGNNAALVYQTSDDSVMVROYGFGNNAATANY 151
DB 104 -----DOWNNAKNSDITVGQYGGNNAALVYQTSDDSVMVROYGFGNNAATANY 151
RESULT 9

AAB36346
ID AAB36346 standard; protein; 151 AA.
XX AAB36346;
AC AAB36346;
DT 26-FEB-2001 (first entry)
XX
XX Agfa::PT3#1 amino acid sequence SEQ ID NO:12.
DE
XX Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
KM vaccine; immune response; immunogen.
XX
XX Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
XX W0200060102-A2.
XX
XX 12-OCT-2000.
PD
XX 05-APR-2000; 2000WO-CA000356.
PF
XX 05-APR-1999; 99US-0127888P.
PR
XX (UYVI-) UNIV VICTORIA.
PA
XX White AP, Doran JL, Collison SK, Kay WW;
PI WPI; 2000-672631/65.
XX N-PSDB; AAC64622.
DR
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 135; 139pp; English.
PS
XX The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAf) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
XX Sequence 151 AA;
SQ
Query Match 78.9%; Score 611; DB 3; Length 151;
Best Local Similarity 80.8%; Pred. No. 3.7e-54;
Matches 122; Conservative 7; Mismatches 22; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVVSGSALAGVVPWGGGNNHNGGNSGDPSTLSTIYQSSANAAALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPWGGGNNHNGGNSGDPSTLSTIYQSSANAAALQ 60

QY 61 SPARKYDQLVTRVTHEMAHAGADNSTIELTONGFRNNAATIDOMNANSDITVGOYGG 120
||| : : : : :
DB 61 SPARKSETTITSGVGNAGDVAGADNSTIELTONGFRNNAATIDOMNANSDITVGOYGG 120
QY 121 NNAALVNOTASDSVMVROVGFNNATANOY 151
||| : : : : :
DB 121 NNAALVNOTASDSVMVROVGFNNATANOY 151
RESULT 10
AAB36347
ID AAB36347 standard; protein; 151 AA.
XX AAB36347;
AC AAB36347;
XX
XX 26-FEB-2001 (first entry)
DT
XX
XX Agfa::PT3#2 amino acid sequence SEQ ID NO:14.
DE
XX Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
KM vaccine; immune response; immunogen.
XX
XX Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
XX W0200060102-A2.
XX
XX 12-OCT-2000.
PD
XX 05-APR-2000; 2000WO-CA000356.
PF
XX 05-APR-1999; 99US-0127888P.
PR
XX (UYVI-) UNIV VICTORIA.
PA
XX White AP, Doran JL, Collison SK, Kay WW;
PI WPI; 2000-672631/65.
XX N-PSDB; AAC64623.
DR
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 136; 139pp; English.
PS
XX The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAf) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in

CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;

Query Match 78.7%; Score 609; DB 3; Length 151;
 Best Local Similarity 81.5%; Pred. No. 5.8e-54;

Matches 123; Conservative 6; Mismatches 22; Indels 0; Gaps 0;

QY 1 MLLKVAAPFAIVSGSALAGVVPWGCGGNHNGSGSPDSTLSIYQGSANALALQ 60

DB 1 MLLKVAAPFAIVSGSALAGVVPWGCGGNHNGSGSPDSTLSIYQGSANALALQ 60

QY 61 SDARKYDQLVTRVVTHEMAHAGQADNSTIELTONGFRNNATIDQNNAKNSDITVQYGG 120

DB 61 SDARKSETTIQSGYNGADVGQADNSTIELTONGFRNNATIDQNNAKNSDITVQYQYDQ 120

QY 121 NNAALVNOTASDSVVRQVFGNNATANQY 151

DB 121 LVTRVVTHEMAHNSVVRQVFGNNATANQY 151

RESULT 11

AAB36352
 ID AAB36352 standard; protein; 151 AA.

AC AAB36352;

DT 26-FEB-2001 (first entry)

DE AgfA:PT3#7 amino acid sequence SEQ ID NO:24.

KW Salmonella; agfA; chromosomal gene replacement; fimbria; epitope;

KM vaccine; immune response; immunogen.

XX Salmonella enteritidis.

OS Escherichia coli.

OS Synthetic.

PN WO200060102-A2.

PD 12-OCT-2000.

PF 05-APR-2000; 2000WO-CA000356.

PR 05-APR-1999; 99US-0127888P.

XX (UUYI-) UNIV VICTORIA.

PI White AP, Doran JL, Collison SK, Kay WM;

DR N-PSDB; AAC64628.

PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence

PT which encodes foreign epitope or antigen, expresses recombinant AgfA

PT protein useful for eliciting immune response in animal.

XX Disclosure; Page 138; 139pp; English.

XX The present invention describes a recombinant agfA gene (1) where a

CC segment of the gene has been replaced by a segment of a foreign DNA

CC sequence which encodes a foreign epitope or antigen. Also described are:

CC (1) use of thin aggregative fimbriae (SEF17/7AP) nucleation depended

CC assembly system of strains of Salmonella, Escherichia coli and

CC Enterobacteriaceae for the production of fimbriae comprising recombinant

CC AgfA, CagA and AgfA-homologue fimbria subunits, respectively; (2)

CC directing recombination of a recombinant gene into the chromosome of the

CC homologous species; (3) directing recombination of a recombinant gene

CC back into the chromosome of the homologous species, replacing the native

CC copy of that gene; and (4) eliciting an immune response in an animal,

CC comprising separating an amino acid polymer comprising a recombinant AgfA

CC protein containing a replacement segment or segments of foreign amino

CC acid sequence or sequences grown on a Salmonella, E. coli or

CC Enterobacteriaceae host cell, from the host cell and introducing the

CC polymer into the animal in conjunction with a carrier or diluent. (1) is

CC useful for the expression of recombinant AgfA protein which is useful for

CC eliciting an immune response in an animal. In a fimbrial presentation

CC system the heterologous antigens are presented in high numbers (up to

CC 500,000 copies/cell), the hybrid fimbria protein possesses both the

CC immunogenicity and adhesion properties relevant for an efficient live

CC vaccine, the carrier fimbrial subunit proteins are usually strong

CC immunogens, which may be important for directing an immune response

CC against the inserted epitope, and hybrid fimbriae are easy and

CC inexpensive to purify in large amount. The present sequence is given in

CC the exemplification of the present invention

XX Sequence 151 AA;

QY 1 MLLKVAAPFAIVSGSALAGVVPWGCGGNHNGSGSPDSTLSIYQGSANALALQ 60

DB 1 MLLKVAAPFAIVSGSALAGVVPWGCGGNHNGSGSPDSTLSIYQGSANALALQ 60

QY 61 SDARKYDQLVTRVVTHEMAHAGQADNSTIELTONGFRNNATIDQNNAKNSDITVQYGG 120

DB 61 SDARKSETTIQSGYNGADVGQADNSTIELTONGFRNNATIDQNNAKNSDITVQYQYDQ 120

QY 121 NNAALVNOTASDSVVRQVFGNNATANQY 151

DB 121 HEMAHANOTASDSVVRQVFGNNATANQY 151

RESULT 12

AAB36355
 ID AAB36355 standard; protein; 151 AA.

AC AAB36355;

DT 26-FEB-2001 (first entry)

DE AgfA:PT3#10 amino acid sequence SEQ ID NO:30.

KW Salmonella; agfA; chromosomal gene replacement; fimbria; epitope;

KM vaccine; immune response; immunogen.

XX Salmonella enteritidis.

OS Escherichia coli.

OS Synthetic.

PN WO200060102-A2.

PD 12-OCT-2000.

PF 05-APR-2000; 2000WO-CA000356.

PR 05-APR-1999; 99US-0127888P.

XX (UUYI-) UNIV VICTORIA.

PI White AP, Doran JL, Collison SK, Kay WM;

DR N-PSDB; AAC64631.

PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence

PT which encodes foreign epitope or antigen, expresses recombinant AgfA

PT protein useful for eliciting immune response in animal.

XX Disclosure; Page 139; 139pp; English.

XX The present invention describes a recombinant agfA gene (1) where a

CC segment of the gene has been replaced by a segment of a foreign DNA

CC sequence which encodes a foreign epitope or antigen. Also described are:

CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC Assembly system of strains of *Salmonella*, *Escherichia coli* and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention

XX Sequence 151 AA;

Query Match 77.1%; Score 597; DB 3; Length 151;
Best Local Similarity 81.5%; Pred. No. 9.8e-53;
Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLKVAAFAAIVVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYQGSANALALQ 60
DB 1 MKLKVAAFAAIVVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYQGSANALALQ 60
QY 61 SDARKYQDLVTRVYTHEMAHNGGADNSTIELTQNGFRNNATTIDQMAKNSDITVGQYGG 120
DB 61 SDARKSETTTTQSGYGAGADVGGADNSTIELTQNGFRNNATTIDQVTRVYTHEMAHNG 120
DB 121 NNAALVNOTASDSSVYVRQVGFNNATANY 151
DB 121 NNAALVNOTASDSSVYVRQVGFNNATANY 151

RESULT 13

AAB36348 standard; protein; 151 AA.

XX AAB36348;
AC AAB36348;
DT 26-FEB-2001 (first entry)
XX Agfa::PT3#3 amino acid sequence SEQ ID NO:16.
DE
XX
KM *Salmonella*; agfa; chromosomal gene replacement; fimbrial epitope;
KM vaccine; immune response; immunogen.
XX
OS *Salmonella enteritidis*.
OS *Escherichia coli*.
OS Synthetic.
XX
XX WO200060102-A2.
XX
XX 12-OCT-2000.
XX
XX 05-APR-2000; 2000WO-CA000356.
XX
XX 05-APR-1999; 99US-0127888P.
XX
XX (UYVI-) UNIV VICTORIA.
XX
XX White AP, Doran JL, Collison SK, Kay WW,
XX
XX WPI, 2000-672631/65.

DR N-PSDB; AAC64624.
XX
PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 136; 139pp; English.

CC The present invention describes a recombinant agfa gene (1) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of *Salmonella*, *Escherichia coli* and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention

XX Sequence 151 AA;

Query Match 74.2%; Score 574; DB 3; Length 151;
Best Local Similarity 80.8%; Pred. No. 2.2e-50;
Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLKVAAFAAIVVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYQGSANALALQ 60
DB 1 MKLKVAAFAAIVVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYQGSANALALQ 60
QY 61 SDARKYQDLVTRVYTHEMAHNGGADNSTIELTQNGFRNNATTIDQMAKNSDITVGQYGG 120
DB 61 SDARKSETTTTQSGYGAGADVGGADNSTIELTQNGFRNNATTIDQMAKNSDITVGQYGG 120
QY 121 NNAALVNOTASDSSVYVRQVGFNNATANY 151
DB 121 NNAALVNOTASDSSVYVRQVGFNNATANY 151

RESULT 14

AAB36343 standard; protein; 151 AA.

XX AAB36343;
AC AAB36343;
DT 26-FEB-2001 (first entry)
XX
XX *Escherichia coli* CsgA amino acid sequence SEQ ID NO:7.
DE
XX
KM *Salmonella*; agfa; chromosomal gene replacement; fimbrial epitope;
KM vaccine; immune response; immunogen.
XX
XX *Escherichia coli*.
XX
XX WO200060102-A2.
XX
XX 12-OCT-2000.

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:25:54 ; Search time 12.4 Seconds
(without alignments)
628.671 Million cell updates/sec

Title: US-09-543-407-20

Perfect score: 774
Sequence: 1 MKLLKVAAPAAIVSGSALA.....DSVWVROVFGNATANQY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/prodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/prodata/2/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/prodata/2/1aa/Backfill1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	684	88.4	151	1 US-08-233-788A-59	Sequence 59, Appl
2	504	65.1	120	1 US-08-233-788A-57	Sequence 57, Appl
3	83.5	10.8	208	4 US-09-252-991A-27661	Sequence 27661, A
4	83.5	10.8	738	3 US-08-864-038A-3	Sequence 3, Appl
5	79.5	10.3	321	4 US-09-498-520A-18	Sequence 18, Appl
6	79.5	10.3	965	1 US-09-328-352-4764	Sequence 4764, Ap
7	78	10.1	363	1 US-08-458-023B-6	Sequence 6, Appl
8	78	10.1	435	2 US-08-331-515A-2	Sequence 2, Appl
9	78	10.1	435	3 US-09-168-406A-2	Sequence 2, Appl
10	77.5	10.0	254	3 US-09-128-450-26	Sequence 26, Appl
11	77.5	10.0	254	4 US-09-823-494-26	Sequence 26, Appl
12	77.5	10.0	364	1 US-07-792-259-17	Sequence 17, Appl
13	77	9.9	943	4 US-09-056-556-204	Sequence 204, App
14	77	9.9	943	4 US-09-072-596-199	Sequence 199, App
15	77	9.9	943	4 US-09-477-135A-131	Sequence 131, App
16	76.5	9.9	943	4 US-09-072-967-204	Sequence 204, App
17	76.5	9.9	1739	4 US-09-540-236-3739	Sequence 3739, Ap
18	76	9.8	352	4 US-09-431-887-32	Sequence 32, Appl
19	76	9.8	273	4 US-09-328-352-6167	Sequence 6167, Ap
20	76	9.8	892	4 US-09-336-447A-5	Sequence 5, Appl
21	76	9.8	2123	3 US-08-968-685A-10	Sequence 10, Appl
22	75.5	9.8	232	4 US-09-555-352-10	Sequence 10, Appl
23	75.5	9.8	254	4 US-09-431-887-23	Sequence 23, Appl
24	75.5	9.8	415	4 US-09-025-769B-280	Sequence 280, App
25	75	9.7	208	3 US-09-128-450-18	Sequence 18, Appl
26	75	9.7	208	4 US-09-823-494-18	Sequence 18, Appl
27	75	9.7	349	4 US-09-300-971A-9	Sequence 9, Appl

28	74.5	9.6	304	1 US-07-851-976B-8	Sequence 8, Appl
29	74.5	9.6	304	1 US-08-291-609-8	Sequence 8, Appl
30	74.5	9.6	304	1 US-08-401-136-8	Sequence 8, Appl
31	74.5	9.6	304	3 US-08-850-554-8	Sequence 8, Appl
32	73.5	9.5	211	1 US-08-276-852-34	Sequence 34, Appl
33	73.5	9.5	211	1 US-08-133-011-16	Sequence 16, Appl
34	73.5	9.5	211	1 US-08-322-730A-16	Sequence 16, Appl
35	73.5	9.5	211	1 US-08-387-874-16	Sequence 16, Appl
36	73.5	9.5	211	1 US-08-899-575-34	Sequence 34, Appl
37	73.5	9.5	211	1 US-08-899-575-34	Sequence 34, Appl
38	73.5	9.5	211	2 US-08-383-619-16	Sequence 16, Appl
39	73.5	9.5	211	3 US-08-907-733-16	Sequence 16, Appl
40	73.5	9.5	211	4 US-09-729-597-16	Sequence 16, Appl
41	73.5	9.5	211	5 PCT-US93-08364-16	Sequence 16, Appl
42	73.5	9.5	211	5 PCT-US95-08743-34	Sequence 34, Appl
43	73.5	9.5	238	4 US-09-495-880A-42	Sequence 42, Appl
44	73.5	9.5	266	4 US-09-495-880A-26	Sequence 26, Appl
45	73.5	9.5	293	3 US-08-438-745-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-08-233-788A-59
; Sequence 59, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, Karen S.
; APPLICANT: Clouthier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,788A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELERX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 59:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 151 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-233-788A-59

Query Match 88.4%; Score 684; DB 1; Length 151;
Best local Similarity 90.1%; Pred. No. 5.5e-66;
Matches 136; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIVSGSALAGVPPWGGGNGNGNSGPPSTISITVYGSANALAIQ 60
|||||

Db 1 MKLKVAFALVVGSLAGVWQGGGNGHNGSGSPDSTLSTIYQGSANALALQ 60
QY 61 SDARKYQDLVTRVVTHEMAHAGADNSTIELTNGFRNNATIDQWNAKSDITVGYG 120
Db 61 SDARKSETTITQSGYGADVGAGADNSTIELTNGFRNNATIDQWNAKSDITVGYG 120
QY 121 NNAALVNOTASDSSVMRQVGFNNATANQY 151
Db 121 NNPALVNOTASDSSVMRQVGFNNATANQY 151

RESULT 2

US-08-233-788A-57
; Sequence 57, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, Karen S.
; APPLICANT: Clouthier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/233, 788A

FILING DATE: 26-APR-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: King, Joshua

REGISTRATION NUMBER: 35,570

REFERENCE/DOCKET NUMBER: 920043, 403C2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

TELEX: 3723836 SREBAMBERY

INFORMATION FOR SEQ ID NO: 57:

SEQUENCE CHARACTERISTICS:

LENGTH: 120 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-233-788A-57

Query Match 65.1%; Score 504; DB 1; Length 120;
Best Local Similarity 87.5%; Pred. No. 9,2e-47;
Matches 98; Conservative 3; Mismatches 11; Indels 0; Gaps 0;
QY 22 VVPWGGGNGHNGSGSPDSTLSTIYQGSANALALQSDARKYDQLVTRVVTHEMAH 81
Db 1 VVPWGGGNGHNGSGSPDSTLSTIYQGSANALALQSDARKSETTITQSGYGADV 60
QY 82 GGGADNSTIELTNGFRNNATIDQWNAKSDITVGYGNNALVNOTASD 133
Db 61 GGGADNSTIELTNGFRNNATIDQWNAKSDITVGYGNNALVNOTASD 112

RESULT 3

US-09-252-991A-27661
; Sequence 27661, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196,136
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27661
; LENGTH: 208
; TYPE: PRN
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27661

Query Match 10.8%; Score 83.5; DB 4; Length 208;
Best Local Similarity 29.0%; Pred. No. 0.35;
Matches 31; Conservative 12; Mismatches 49; Indels 15; Gaps 3;

QY 3 LKVAFAALVVGSL-----AGVPEWGGGNGHNGSGSPDST-----LSLY 48
Db 101 LFAVAAL-GILFAGSTLPGELSPVSAAYAKGGNGGSGHSGKGGKGNLGHSSK 159
QY 49 QYGSANALALQSDARKYDQLVTRVVTHEMAHAGADNSTIELTNG 95
Db 160 GHOSATSGIASRDSRGLSQASAIATTPGDHNSKLSNAIGSSYKN 206

RESULT 4

US-08-864-038A-3
; Sequence 3, Application US/08864038A
; Patent No. 6001592

GENERAL INFORMATION:

APPLICANT: Kunio NAKASHIMA et al.

TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE CDNA, VECTOR

TITLE OF INVENTION: CONTAINING SAID CDNA, HOST CELLS TRANSFORMED WITH SAID

TITLE OF INVENTION: VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING

TITLE OF INVENTION: SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBODY

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: 812-5 Hirano

STREET: Ishinden

CITY: Tau-city

STATE: Mie-prefecture

COUNTRY: JAPAN

ZIP: 514-01

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage

COMPUTER: IBM Compatible

OPERATING SYSTEM: Microsoft Windows 95

SOFTWARE: Word Perfect 6.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/864, 038A

FILING DATE: May 28, 1997

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 8-184459

FILING DATE: 15-July-1996

ATTORNEY/AGENT INFORMATION:

NAME: C. Bruce Hamburg

REGISTRATION NUMBER: 22,389

REFERENCE/DOCKET NUMBER: F-5610

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 966-2340

TELEFAX: (212) 953-7733

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 738

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

ORIGINAL SOURCE:
ORGANISM: Pinctada fucata
CELL TYPE: mantle epithelial cell
FEATURE:
NAME/KEY: peptide
LOCATION: from 1 to 738
IDENTIFICATION METHOD: E (by experiment)
US-08-864-038A-3

Query Match 10.8%; Score 83.5; DB 3; Length 738;
Best Local Similarity 27.7%; Pred. No. 2;
Matches 44; Conservative 12; Mismatches 56; Indels 47; Gaps 6;

QY 3 LKVAAPAAIVVSGSALAGVPMWGGGNNHGGNSGPDSTLSIYQGSANALALQSD 62
DB 419 LKSSASASASASASAG-----GGGGGGGGGGGGGG-----GGGALALALA 465
QY 63 AKRYDQ-----VTRVTHEMAHAGGADNSTIELTONGFRNATTIDQNAKSDITVQY 118
DB 466 AAGGGGAGGGGGGALALALAAAGG-----GGGGLGGL-----GG 504
QY 119 GGNNAALVNOTASDS-----VMVROVFGNNATA 148
DB 505 GGGSAALAAAAAAGGGGGRALRLRLRMRGGSSAAA 543

RESULT 5

US-09-498-520A-18
Sequence 18, Application US/09498520A
Patent No. 6613553

GENERAL INFORMATION:
APPLICANT: Rock, Charles O
APPLICANT: Heath, Richard J
TITLE OF INVENTION: No. 6613553el Enoyl Reductases and Methods of Use Thereof
FILE REFERENCE: SJ-0022
CURRENT APPLICATION NUMBER: US/09/498,520A
CURRENT FILING DATE: 2000-02-04
NUMBER OF SEQ ID NOS: 62
SOFTWARE: PatentIn version 3.1
SEQ ID NO 18
LENGTH: 321
TYPE: PRT
ORGANISM: Caulobacter crescentus
US-09-498-520A-18

Query Match 10.3%; Score 79.5; DB 4; Length 321;
Best Local Similarity 24.5%; Pred. No. 1.7;
Matches 48; Conservative 15; Mismatches 64; Indels 69; Gaps 10;

QY 4 LKVAAPAAIVVSGS-----LAG-----VPMWGGGNNHNG----- 34
DB 109 LKAAAGKAVVCGAVGAVKAEAGCDAYICGGEGGHTGLVTEPLVAQAVEANKIV 168
QY 35 ---GNSGPDSTLSIYQGSANALALQSDARKYDQVTRVTHEMANG-----Q 83
DB 169 VAAGGHHG-----RGLA--ALALAGQG-----VMGTRPLASHEAHAGDLVROAVVE 215
QY 84 GADNSTIEL-TONG-----FRNATTIDQNAKSDITV-----QYGNNAALVN 127
DB 216 AADSDTVTRCYSGKPMRYKKNPYDWEARPDIDPFOQAVNISIRNGMGGISGQIEG 275
QY 128 QTSADSSVMVROYGFG 143
DB 276 LDAKSCFAMGOSAGC 291

RESULT 6

US-09-328-352-4764
Sequence 4764, Application US/09328352
Patent No. 6562958

GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER

TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328,352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 4764
LENGTH: 975
TYPE: PRT
ORGANISM: Acinetobacter baumannii
US-09-328-352-4764

Query Match 10.3%; Score 79.5; DB 4; Length 975;
Best Local Similarity 25.0%; Pred. No. 8;
Matches 34; Conservative 19; Mismatches 70; Indels 13; Gaps 4;

QY 15 SGALAGVPMWGGGNNHGGNSGPDSTLSIYQGSANALALQSDARKYDQVTRV 74
DB 279 NGTGSBG-VSALGGSGSGDAGNGIASGNGENHYIGNG-----NGDDVDITAPITGV 333
QY 75 THE-----MAHAGGADNSTIELTONGFRNATTIDQNAKSDITVQYGNNAALVNQ 128
DB 334 NISGNSFTLIGNSSSVTAPTTSTNVNDITD--NGNGGNGSGSGSGDGLNG 391
QY 129 TMSDSSVMVROYGFGN 144
DB 392 AASGNGENHYIGNGN 407

RESULT 7

US-08-458-023B-6
Sequence 6, Application US/08458023B
Patent No. 5667990

GENERAL INFORMATION:
APPLICANT: Berka, Randy M.
APPLICANT: Yoder, Wendy
APPLICANT: Takagi, Shinobu
APPLICANT: Boomatinan, Karuppan C.
TITLE OF INVENTION: ASPERGILLUS EXPRESSION SYSTEM
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSER: No. 56679900 No. 56679900disk of No. 56679900th America, Inc.
STREET: 405 Lexington Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10174-6201

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/458,023B
FILING DATE: 01-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Lowney Dr., Karen A.
REGISTRATION NUMBER: 31,274
TELEPHONE: 212-867-0123
TELECOMMUNICATION INFORMATION:
REFERENCE/DOCKET NUMBER: 4086.010-US

INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 363 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-458-023B-6

Query Match 10.1%; Score 78; DB 1; Length 363;
Best Local Similarity 30.7%; Pred. No. 2.9;
Matches 35; Conservative 13; Mismatches 28; Indels 38; Gaps 8;

QY 1 MKLKVAFAFAIVSGSALAGVVPWGGGNNH---NGGNSSG-----PDSTLSI 47
 DB 1 MKLSLSTFAAVITIGALAL---PQGGGGSVTCPPGGGTSNSGCCWCFDVLDDLTQNF 56
 QY 48 YQGSANALALQSDARKYDQLVTRVYTHE-----MAHAQO-----GADNSTI 90
 DB 57 YQ-GS-----KCESPVK-----ILRIVFHDAIGFSPALTAAGQFGGGADGSI 100

RESULT 8
 US-08-331-515A-2
 ; Sequence 2, Application US/08331515A
 ; Patent No. 5851811
 ; GENERAL INFORMATION:
 ; APPLICANT: Melinder, Karen
 ; APPLICANT: Andersen, Morten B
 ; TITLE OF INVENTION: PEROXIDASE VARIANTS WITH IMPROVED
 ; NUMBER OF SEQUENCES: 2
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: No. 58518110 No. 5851811disk of No. 5851811th America, Inc.
 ; STREET: 405 Lexington Avenue, 64th Floor
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: United States of America
 ; ZIP: 10174-6401
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/331.515A
 ; FILING DATE: 01-NOV-1994
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Harrington, James J.
 ; REGISTRATION NUMBER: 38,711
 ; REFERENCE/DOCKET NUMBER: 3769.204-US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 212-867-0123
 ; TELEFAX: 212-878-9655
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 435 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-08-331-515A-2

Query Match 10.1%; Score 78; DB 2; Length 435;
 Best Local Similarity 30.7%; Pred. No. 3.8;
 Matches 35; Conservative 13; Mismatches 28; Indels 38; Gaps 8;

QY 1 MKLKVAFAFAIVSGSALAGVVPWGGGNNH---NGGNSSG-----PDSTLSI 47
 DB 2 MKLSLSTFAAVITIGALAL---PQGGGGSVTCPPGGGTSNSGCCWCFDVLDDLTQNF 57
 QY 48 YQGSANALALQSDARKYDQLVTRVYTHE-----MAHAQO-----GADNSTI 90
 DB 58 YQ-GS-----KCESPVK-----ILRIVFHDAIGFSPALTAAGQFGGGADGSI 101

RESULT 9
 US-09-168-406A-2
 ; Sequence 2, Application US/09168406A
 ; Patent No. 6258769
 ; GENERAL INFORMATION:
 ; APPLICANT: Melinder, Karen G.
 ; APPLICANT: Andersen, Morten B.
 ; TITLE OF INVENTION: Peroxidase Variants With Improved
 ; TITLE OF INVENTION: Hydrogen Peroxidase Stability

FILE REFERENCE: 3769.214-US
 ; CURRENT APPLICATION NUMBER: US/09/168,406A
 ; CURRENT FILING DATE: 1998-10-06
 ; PRIOR APPLICATION NUMBER: 08/331,515
 ; PRIOR FILING DATE: 1994-11-01
 ; PRIOR APPLICATION NUMBER: PCT/DK93/00189
 ; PRIOR FILING DATE: 1993-06-01
 ; PRIOR APPLICATION NUMBER: 0792/92
 ; PRIOR FILING DATE: 1992-06-01
 ; NUMBER OF SEQ ID NOS: 43
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 2
 ; LENGTH: 435
 ; TYPE: PRT
 ; ORGANISM: Coprinus cinereus
 ; FEATURE:
 ; NAME/KEY: VARIANT
 ; LOCATION: (1)...(435)
 ; OTHER INFORMATION: Xaa = Any Amino Acid
 ; US-09-168-406A-2

Query Match 10.1%; Score 78; DB 3; Length 435;
 Best Local Similarity 30.7%; Pred. No. 3.8;
 Matches 35; Conservative 13; Mismatches 28; Indels 38; Gaps 8;

QY 1 MKLKVAFAFAIVSGSALAGVVPWGGGNNH---NGGNSSG-----PDSTLSI 47
 DB 2 MKLSLSTFAAVITIGALAL---PQGGGGSVTCPPGGGTSNSGCCWCFDVLDDLTQNF 57
 QY 48 YQGSANALALQSDARKYDQLVTRVYTHE-----MAHAQO-----GADNSTI 90
 DB 58 YQ-GS-----KCESPVK-----ILRIVFHDAIGFSPALTAAGQFGGGADGSI 101

RESULT 10
 US-09-128-450-26
 ; Sequence 26, Application US/09128450
 ; Patent No. 6211149
 ; GENERAL INFORMATION:
 ; APPLICANT: Cheesbro, Bruce W
 ; APPLICANT: Caughey, Byron W
 ; APPLICANT: Chabry, Joelle
 ; APPLICANT: Priola, Susette
 ; TITLE OF INVENTION: Inhibitors of Formation of Protease Resistant Prion
 ; FILE REFERENCE: 50121
 ; CURRENT APPLICATION NUMBER: US/09/128,450
 ; CURRENT FILING DATE: 1998-08-03
 ; NUMBER OF SEQ ID NOS: 29
 ; SOFTWARE: Patentin Ver. 2.0
 ; SEQ ID NO 26
 ; LENGTH: 254
 ; TYPE: PRT
 ; ORGANISM: Hamster sp.
 ; US-09-128-450-26

Query Match 10.0%; Score 77.5; DB 3; Length 254;
 Best Local Similarity 25.4%; Pred. No. 2;
 Matches 32; Conservative 19; Mismatches 60; Indels 15; Gaps 6;

QY 26 WG-GGNNHGGGSGSDSTLSIYQGSANALALQSDARKYDQLVTRVYTHEMAHAQO 84
 DB 89 WGCGGTHNOWNKSPKRYNMK--HMAAAGAAVVGGLGY--MLGSAMSRPMHFGND 144
 QY 85 ADNSTIELTQNGFNNNA---TIIDMAKNS-----DIYGYGGANNAI-VNQTASDS 134
 DB 145 WEDRYRENMRYPNOYVPRVDQYNNNNFVHDCVNTIKKHVTVTTTTGTGENFTETDIK 204
 QY 135 VNVQV 140
 DB 205 IWERVV 210

RESULT 11
US-09-823-494-26
Sequence 26, Application US/09823494
Patent No. 6355610
GENERAL INFORMATION:
APPLICANT: Chesebro, Bruce W
APPLICANT: Caughey, Byron W
APPLICANT: Chabry, Joelle
APPLICANT: Pricola, Suelette
TITLE OF INVENTION: Inhibitors of Formation of Protease Resistant Prion
TITLE OF INVENTION: Protein
FILE REFERENCE: 50121
CURRENT APPLICATION NUMBER: US/09/823,494
CURRENT FILING DATE: 2001-03-30
PRIOR APPLICATION NUMBER: 09/128,450
PRIOR FILING DATE: 1998-08-03
NUMBER OF SEQ ID NOS: 29
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 26
LENGTH: 254
TYPE: PRT
ORGANISM: Hamster sp.
US-09-823-494-26
Query Match 10.0%; Score 77.5; DB 4; Length 254;
Best Local Similarity 25.4%; Pred. No. 2;
Matches 32; Conservative 19; Mismatches 60; Indels 15; Gaps 6;
QY 26 WG-GGNNHGGGSSGPDSTLSIYQGSANALALOSDARKYDQVTVRYTHMAHAGG 84
DB 89 WGGGGGTHQWKNPKSPKTKMK--HMAGAAAGAVVGGGGLGY--MLGSANSRPMHFGND 144
QY 85 ADNSTELTONGFRNNA---TIDQWNAKNS-----DITVQYGGNNAAL-VNQTASDSS 134
DB 145 WEDRYRRENNRRPNQVYRPVDQYNNQNNFVADCNITTKQHTVTTTGTGKRFETDIX 204
QY 135 VNVQVY 140
DB 205 IMERVV 210
RESULT 12
US-07-792-259-17
Sequence 17, Application US/07792259
Patent No. 5286638
GENERAL INFORMATION:
APPLICANT: TANAKA, YOSHIKAZU
APPLICANT: ASHIKARI, TOSHIHIKO
APPLICANT: HATANAKA, HARUYO
APPLICANT: SHIBANO, YUJI
APPLICANT: AMACHI, TERUO
APPLICANT: NAKAYAMA, TORU
APPLICANT: SUMIDA, MOTOO
TITLE OF INVENTION: PEROXIDE GENE OF MICROBIAL ORIGIN
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSES: CUSHMAN, DARBY & CUSHMAN
STREET: 1615 L STREET, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: USA
ZIP: 20036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/792,259
FILING DATE: 19911115
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: SCOTT, WATSON T.

REGISTRATION NUMBER: 26581
REFERENCE/DOCKET NUMBER: 9437/93433
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-861-3067
TELEFAX: 202-892-0944
TELEX: 6714627 CUSH
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 364 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULAR TYPE: peptide
US-07-792-259-17
Query Match 10.0%; Score 77.5; DB 1; Length 364;
Best Local Similarity 30.4%; Pred. No. 3.3;
Matches 35; Conservative 13; Mismatches 28; Indels 39; Gaps 8;
QY 1 MKLKVAFAPAVIVSGSALAGVVPQWGGGNNH---GGGNSG-----PDSTLS 46
DB 1 MKLSLSTFAVITIGALAL---PQGGGGGSGVTCFGGQSTNSGCCVWFDVLDLQTN 56
QY 47 IYQGSANALALOSDARKYDQVTVRYTHMAHAGG-----GADNSTI 90
DB 57 FYQ-GS-----KCESPVRK-----ILRIYFHDAIGSPALTAAGQFGGAGDSII 101
RESULT 13
US-09-056-556-204
Sequence 204, Application US/09056556
Patent No. 6350456
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, David C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
NUMBER OF SEQUENCES: 241
CORRESPONDENCE ADDRESS:
ADDRESSES: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/056,556
FILING DATE: 07-APR-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.457
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 204:
SEQUENCE CHARACTERISTICS:
LENGTH: 943 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-09-056-556-204
Query Match 9.9%; Score 77; DB 4; Length 943;
Best Local Similarity 26.2%; Pred. No. 14;
Matches 37; Conservative 14; Mismatches 66; Indels 24; Gaps 7;

QY 16 GSALAGVVPQWGG-GNHN-GGNSGSPDSTLSTIYQGSANAA--LALQSDARKYDQVLT 71
 DB 464 GSGNIGVFNVGSGSLGNVNIISGN-----LGIYNIIGFNVGDYVNVFGNAGDPNQGFA 516
 QY 72 RVVTHEMAHAGGADNSTIELT---QNGFRNNATIDQNNAKNSDITVQYGGNNAALVNO 128
 DB 517 NTGNNNIGFANTGNNNIGIGLSGDNQGFN---IASGWNSTGNSGLFNSGTNNVGIFNA 573
 QY 129 TASDSSVWVRQYGFNNATAN 149
 DB 574 GTGN-----VGIANSTGN 587

RESULT 14
 US-09-072-596-199
 ; Sequence 199; Application US/09072596
 ; Patent No. 6458366
 ; GENERAL INFORMATION:
 ; APPLICANT: Reed, Steven G.
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Campos-Neto, Antonia
 ; APPLICANT: Houghton, Raymond
 ; APPLICANT: Vardick, Thomas S.
 ; APPLICANT: Twardzik, Daniel R.
 ; APPLICANT: Lodes, Michael J.
 ; APPLICANT: Hendrickson, Ronald C.
 ; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
 ; NUMBER OF SEQUENCES: 350
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SEED and BERRY LLP
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue
 ; CITY: Seattle
 ; STATE: Washington
 ; COUNTRY: USA
 ; ZIP: 98104-7092
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentln Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/072.596
 ; FILING DATE: 05-MAY-1998
 ; CLASSIFICATION:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Makl, David J.
 ; REGISTRATION NUMBER: 31,392
 ; REFERENCE/DOCKET NUMBER: 210121.417C9
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (206) 622-4900
 ; TELEFAX: (206) 682-6031
 ; INFORMATION FOR SEQ ID NO: 199:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 943 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS:
 ; TOPOLOGY: linear
 ; US-09-072-596-199

Query Match 9.9%; Score 77; DB 4; Length 943;
 Best Local Similarity 26.2%; Pred. No. 14;
 Matches 37; Conservative 14; Mismatches 66; Indels 24; Gaps 7;

QY 16 GSALAGVVPQWGG-GNHN-GGNSGSPDSTLSTIYQGSANAA--LALQSDARKYDQVLT 71
 DB 464 GSGNIGVFNVGSGSLGNVNIISGN-----LGIYNIIGFNVGDYVNVFGNAGDPNQGFA 516
 QY 72 RVVTHEMAHAGGADNSTIELT---QNGFRNNATIDQNNAKNSDITVQYGGNNAALVNO 128
 DB 517 NTGNNNIGFANTGNNNIGIGLSGDNQGFN---IASGWNSTGNSGLFNSGTNNVGIFNA 573
 QY 129 TASDSSVWVRQYGFNNATAN 149

DB 574 GTGN-----VGIANSTGN 587
 RESULT 15
 US-09-477-135A-131
 ; Sequence 131; Application US/09477135A
 ; Patent No. 6572865
 ; GENERAL INFORMATION:
 ; APPLICANT: Nano, Francis
 ; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
 ; TITLE OF INVENTION: Immunostimulatory Peptides
 ; FILE REFERENCE: 52868
 ; CURRENT APPLICATION NUMBER: US/09/477,135A
 ; CURRENT FILING DATE: 2000-01-03
 ; PRIOR APPLICATION NUMBER: 08990823
 ; PRIOR FILING DATE: 1997-12-15
 ; PRIOR APPLICATION NUMBER: US 96/10375
 ; PRIOR FILING DATE: 1996-06-14
 ; PRIOR APPLICATION NUMBER: 60/000,254
 ; PRIOR FILING DATE: 1995-06-15
 ; NUMBER OF SEQ ID NOS: 169
 ; SOFTWARE: Patentln Ver. 2.0
 ; SEQ ID NO 131
 ; LENGTH: 943
 ; TYPE: PRT
 ; ORGANISM: Mycobacterium tuberculosis
 ; US-09-477-135A-131

Query Match 9.9%; Score 77; DB 4; Length 943;
 Best Local Similarity 26.2%; Pred. No. 14;
 Matches 37; Conservative 14; Mismatches 66; Indels 24; Gaps 7;
 QY 16 GSALAGVVPQWGG-GNHN-GGNSGSPDSTLSTIYQGSANAA--LALQSDARKYDQVLT 71
 DB 574 GSGNIGVFNVGSGSLGNVNIISGN-----LGIYNIIGFNVGDYVNVFGNAGDPNQGFA 623
 QY 72 RVVTHEMAHAGGADNSTIELT---QNGFRNNATIDQNNAKNSDITVQYGGNNAALVNO 128
 DB 624 NTGNNNIGFANTGNNNIGIGLSGDNQGFN---IASGWNSTGNSGLFNSGTNNVGIFNA 680
 QY 129 TASDSSVWVRQYGFNNATAN 149
 DB 681 GTGN-----VGIANSTGN 694

Search completed: March 11, 2004, 18:44:51
 Job time : 12.4 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Comogen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:40:56 ; Search time 24.6 Seconds

(without alignments)
1296.103 Million cell updates/sec

Title: US-09-543-407-20

Perfect score: 774
Sequence: 1 MRLKVAAPAAIVSGSALA.....DSVMYROYFGNNATANY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 809742 seqs, 21153259 residues

Total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.*

```
1: /cgn2_6/ptodata/1/pubppaa/PCT_NEW_PUB.pep.*
2: /cgn2_6/ptodata/1/pubppaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubppaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubppaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubppaa/PCT_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubppaa/PCT_NEW_PUB.pep.*
7: /cgn2_6/ptodata/1/pubppaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubppaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubppaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubppaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubppaa/US09_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubppaa/US09_PUBCOMB.pep.*
13: /cgn2_6/ptodata/1/pubppaa/US10_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubppaa/US10_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubppaa/US10_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubppaa/US10_PUBCOMB.pep.*
17: /cgn2_6/ptodata/1/pubppaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubppaa/US60_PUBCOMB.pep.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	101.5	13.1	445	US-10-369-493-20638	Sequence 20638, A
2	94.5	12.2	438	US-10-156-761-9343	Sequence 9343, Ap
3	84.5	10.9	1778	US-10-238-075-749	Sequence 749, App
4	83.5	10.8	278	US-09-810-264-28	Sequence 28, App1
5	83	10.7	597	US-09-793-306-146	Sequence 146, App
6	82.5	10.7	435	US-10-128-714-3213	Sequence 3213, Ap
7	82.5	10.7	515	US-10-128-714-8213	Sequence 8213, Ap
8	82.5	10.7	2732	US-10-238-075-1119	Sequence 1119, App
9	82.5	10.7	2834	US-10-085-959-252	Sequence 252, App
10	82	10.6	65	US-09-996-194-16	Sequence 16, App1
11	82	10.6	354	US-09-820-843A-21	Sequence 21, App1
12	80.5	10.4	486	US-10-369-493-20619	Sequence 20619, A
13	79.5	10.3	562	US-10-156-761-13039	Sequence 13039, A
14	79	10.2	276	US-10-369-493-3641	Sequence 3641, Ap
15	79	10.2	688	US-10-032-585-7876	Sequence 7876, Ap

16	78.5	10.1	271	US-10-156-761-11721	Sequence 11721, A
17	78	10.1	545	US-10-369-493-18473	Sequence 18473, A
18	78	10.1	594	US-10-156-761-13173	Sequence 13173, A
19	77.5	10.0	254	US-10-410-907A-5	Sequence 5, App1
20	77	9.9	943	US-09-996-634-131	Sequence 131, App
21	77	9.9	943	US-09-997-182-131	Sequence 131, App
22	77	9.9	943	US-09-997-181-131	Sequence 131, App
23	77	9.9	943	US-10-193-002-199	Sequence 199, App
24	77	9.9	943	US-10-084-843-204	Sequence 204, App
25	76.5	9.9	254	US-10-301-488A-26	Sequence 32, App1
26	76	9.8	852	US-10-304-630-32	Sequence 32, App1
27	76	9.8	892	US-09-952-257-5	Sequence 9, App1
28	76	9.8	2122	US-09-813-214A-9	Sequence 23, App1
29	75.5	9.8	254	US-10-304-630-23	Sequence 23, App1
30	75.5	9.8	254	US-10-410-907A-7	Sequence 7, App1
31	75.5	9.8	482	US-10-156-761-8763	Sequence 8763, Ap
32	75.5	9.8	552	US-10-369-493-1745	Sequence 3745, Ap
33	75.5	9.8	2893	US-09-882-227-522	Sequence 522, App
34	75	9.7	250	US-08-779-457-49	Sequence 49, App1
35	75	9.7	254	US-09-880-748-1186	Sequence 1186, Ap
36	75	9.7	358	US-10-156-761-9634	Sequence 9634, Ap
37	74	9.6	253	US-09-880-748-2098	Sequence 2098, Ap
38	74	9.6	284	US-10-156-761-8340	Sequence 8340, Ap
39	74	9.6	350	US-10-156-761-13168	Sequence 13168, A
40	74	9.6	400	US-10-009-823A-10	Sequence 10, App1
41	74	9.6	402	US-09-712-363-147	Sequence 147, App
42	73.5	9.5	211	US-10-016-986-14	Sequence 34, App1
43	73.5	9.5	211	US-10-273-973-16	Sequence 16, App1
44	73.5	9.5	219	US-09-809-517A-24	Sequence 24, App1
45	73.5	9.5	219	US-09-809-517A-27	Sequence 27, App1

ALIGNMENTS

```
RESULT 1
US-10-369-493-20638
; Sequence 20638, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052) B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 20638
; LENGTH: 445
; TYPE: PRT
; ORGANISM: Rhodospseudomonas palustris
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(445)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-20638
```

Best Match 13.1%: Score 101.5; DB 15; Length 445;

Query Local Similarity 26.0%: Pred No. 0.046; Mismatches 55; Indels 53; Gaps 7;

```
QY 7 AAFAA-----IVSGSALAGVPPQGGG-----NHNGG-----GNSSGPDSTLSIYQ 50
|||
DB 19 AAFADSNVTYVYINQGNDDQANITQSGNNGSVGAFNGSGFLQENGTLISGA-NLLTVKQS 77
|||
QY 51 GSANALALQSDARKYDQYLVTRVTVHEMAHQGA-----DNSTIELQNGFRNNATID 104
|||
```

Db 78 GNSNS-----VGRDIQKQSGAGNSAII FQEGTSGDVELQOTGTSGAVPS 123

Qy 105 OMNAKN-----SDITVGYGGNNALVNOTASDSSVWROYG 141
 Db 124 GMMNTNPGVFNKITTODSSNGSKSVYIOPGKNVFSIKQGMTGNTSTVNOIG 176

RESULT 2

US-10-156-761-9343
 ; Sequence 9343, Application US/10156761
 ; Publication No. US20030119018A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OMURA, SATOSHI
 ; APPLICANT: IKEDA, HARUO
 ; APPLICANT: ISHIGAWA, JUN
 ; APPLICANT: HORIKAWA, HIROSHI
 ; APPLICANT: SHIBA, TADAYOSHI
 ; APPLICANT: SAKAKI, YOSHIYUKI
 ; APPLICANT: HATTORI, MASAHIRA
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-262
 ; CURRENT APPLICATION NUMBER: US/10/156,761
 ; PRIOR FILING DATE: 2002-05-29
 ; PRIOR APPLICATION NUMBER: JP 2001-204089
 ; PRIOR FILING DATE: 2001-05-30
 ; PRIOR APPLICATION NUMBER: JP 2001-272697
 ; NUMBER OF SEQ ID NOS: 15109
 ; SEQ ID NO 9343
 ; LENGTH: 438
 ; TYPE: PRN
 ; ORGANISM: Streptomyces avermitilis
 US-10-156-761-9343

Query Match 12.2%; Score 94.5; DB 14; Length 438;
 Best Local Similarity 25.9%; Pred. No. 0.24;
 Matches 37; Conservative 27; Mismatches 46; Indels 33; Gaps 8;

Qy 1 MKLTKAAPFAIVVSSSALAGVVPWGGGNNHGGSSGPDSTLSIYOGSANAALA-- 58
 Db 1 MRSIRAAAGVAVTMSIALAASAC---GGGSGTGGGSDSP-KTLT-YMASNOGASIVAD 54
 Qy 59 ---LQSDARKYDO-----LVTRVVTHEMAHAGGADNSTELTON-GFRN 99
 Db 55 KKVLEPELDKFEEOGICIKVLEVPWSDILNRILT--ATTSGGQPDVLTINGTWMSLSQA 112
 Qy 100 NATIDOMNAKNSDITVGYGGNN 122
 Db 113 TGAALPMDAKNFD---KIGGKD 131

RESULT 3

US-10-238-075-749
 ; Sequence 749, Application US/10238075
 ; Publication No. US20030148324A1
 ; GENERAL INFORMATION:
 ; APPLICANT: I.N.S.E.R.M.
 ; TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isolated
 ; FILE REFERENCE: BLANDINE
 ; CURRENT APPLICATION NUMBER: US/10/238,075
 ; PRIOR FILING DATE: 2002-09-10
 ; PRIOR APPLICATION NUMBER: 0003145
 ; NUMBER OF SEQ ID NOS: 1576
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 749
 ; LENGTH: 1778
 ; TYPE: PRN
 ; ORGANISM: Escherichia coli
 US-10-238-075-749

Query Match 10.9%; Score 84.5; DB 14; Length 1778;

Best Local Similarity 26.8%; Pred. No. 15;
 Matches 40; Conservative 21; Mismatches 59; Indels 29; Gaps 7;

Qy 5 KVAAPFAIVVSSSALAGVVPW---GGGNNHGGSSGPDSTLSIYOGSANAALA-- 59
 Db 906 KISSNSTDANGSQLYVADSFTSYLGGGADISDTVLSGPTTYIGCTTYTWGDLAAI 965
 Qy 60 -----QSDARKYDQLVTRVVTHEMAHAGGADNSTELTO--NGFRNATIDOMNAKN 110
 Db 966 NTSEFSTLGDALLMD-----ATAGKFSAGHGNNAPSVITDVANGAVSTSSDAINGSQ 1019
 Qy 111 ---SDITVGYGGNNALVNOTASDSSV 135
 Db 1020 LYGVSDIYDALGN--AVVN--TDGSI 1043

RESULT 4

US-09-810-264-28
 ; Sequence 28, Application US/09810264
 ; Patent No. US2002007675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Crane, Virginia C.
 ; APPLICANT: Famodu, Omolayo O.
 ; APPLICANT: Hu, Xu
 ; APPLICANT: Lu, Guihua
 ; APPLICANT: Zhang, Lingyu
 ; TITLE OF INVENTION: WRKY Transcription Factors and Methods
 ; FILE REFERENCE: 1183
 ; CURRENT APPLICATION NUMBER: US/09/810,264
 ; PRIOR FILING DATE: 2001-03-16
 ; PRIOR APPLICATION NUMBER: US 60/190,467
 ; NUMBER OF SEQ ID NOS: 43
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 28
 ; LENGTH: 278
 ; TYPE: PRN
 ; ORGANISM: Trifolium aestivum
 US-09-810-264-28

Query Match 10.8%; Score 83.5; DB 9; Length 278;
 Best Local Similarity 20.8%; Pred. No. 1.8;
 Matches 37; Conservative 23; Mismatches 51; Indels 67; Gaps 6;

Qy 15 SSSALAGVVPWGGGNNHGGSSGPDSTLSIYOGSANAALQSDARKYDQLVTRRV 74
 Db 71 NSSGCAAVTAE-----DHTNGSEHSGPPENSVTFGDDADNGAEPETKRKE----- 119
 Qy 75 TH---EMAHAGGA-----DNSTLELTONGFR----- 98
 Db 120 -HQDNSSGGTGACVVRVREPRVLVOTLSDIDLDGFFWRKRYGQKVVGKGNPNPSYRK 178
 Qy 99 -----NNAITIDOMNAKNS-DITVGYGGNNALVNOTASDSSVMV 137
 Db 179 CTTVCQVPRKVERASHDNRAVITTYTGRSHSDVPPVGRGAGSRALPTSSSDSSVYV 236

RESULT 5

US-09-793-306-146
 ; Sequence 146, Application US/09793306
 ; Patent No. US20020098200A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Campos-Neto, Antonio
 ; APPLICANT: Skeiky, Yasir
 ; APPLICANT: Ovendale, Pamela
 ; APPLICANT: Jen, Shyian
 ; APPLICANT: Lodes, Michael
 ; APPLICANT: Corixa Corporation
 ; TITLE OF INVENTION: Compounds and Methods for Diagnosis and Immunotherapy
 ; FILE REFERENCE: 014058-008740US
 ; CURRENT APPLICATION NUMBER: US/09/793,306

QY 70 ---VTRVTHMAHAGQ--GADNSTIELTNGFR-----NNATIDQ-----WN 107
DB 330 TDFALTAAYHD--HTGKVPESDTEKVAVAKKREFOAMVNNKTVKELVASADDFMTYIMD 387
QY 108 AKNSDITVGQYGGNNALVNTQASDSSVMVROYGFQNN 145
DB 388 PENSTKPIARILGHQKEVNVHTFSPDMAYIASAGFDNH 425
RESULT 7
US-10-128-714-8213
; Sequence 8213, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wengqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Broshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8213
; LENGTH: 515
; TYPE: PRT
; ORGANISM: Aspergillus fumigatus
US-10-128-714-8213
Query Match 10.7%; Score 82.5; DB 14; Length 515;
Best Local Similarity 19.6%; Pred. No. 5;
Matches 31; Conservative 30; Mismatches 66; Indels 31; Gaps 6;
QY 13 VVSGSALAGVPOWGGGNNHNGGNSGPDSTSIYQSGANALALQSDARKYDQ--- 69
DB 274 VLTGHGSGVTCVWAGGTGKI-----YTSSHDRTIKIMWAGNSLIQTLSAHNRVNHILS 329
QY 70 ---VTRVTHMAHAGQ--GADNSTIELTNGFR-----NNATIDQ-----WN 107
DB 330 TDFALTAAYHD--HTGKVPESDTEKVAVAKKREFOAMVNNKTVKELVASADDFMTYIMD 387
QY 108 AKNSDITVGQYGGNNALVNTQASDSSVMVROYGFQNN 145
DB 388 PENSTKPIARILGHQKEVNVHTFSPDMAYIASAGFDNH 425
RESULT 8
US-10-238-075-1119
; Sequence 1119, Application US/10238075
; Publication No. US20030148324A1
; GENERAL INFORMATION:
; APPLICANT: I.N.S.E.R.M.
; TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isolated
; FILE REFERENCE: BLANDINE
; CURRENT APPLICATION NUMBER: US/10/238,075
; CURRENT FILING DATE: 2002-09-10
; PRIOR APPLICATION NUMBER: 0003145

PRIOR FILING DATE: 2000-03-10
 NUMBER OF SEQ ID NOS: 1576
 SOFTWARE: Patentin version 3.1
 SEQ ID NO: 1119
 LENGTH: 2732
 TYPE: PRT
 ORGANISM: Escherichia coli
 US-10-238-075-1119

Query Match 10.7%; Score 82.5; DB 14; Length 2732;
 Best Local Similarity 27.8%; Pred. No. 43;
 Matches 42; Conservative 16; Mismatches 54; Indels 39; Gaps 7;

QY 8 AFAIVSGSALAGVVPQWGG---GNHGGN---SSGPDSTIYQGSANA--ALA 58
 DB 283 ASGRITVNNSLATGAVTAKGQVTLTGPHKAGNLSVSRDIVLS---NGTLNSDKDLS 339
 QY 59 LOSDARKYDQVTRVVTHEMAHAGAGADNSTIELTNGFRNNATIDQWNAKNSDITVGY 118
 DB 340 LTAGGRITQ-----NEKLTAGRDVTLAAKNITD-----TASQINAAADIVTV--- 383
 QY 119 GGNNAALVNOTASDSSVWVROVFGNNATAN 149
 DB 384 -----ASDITLTGQITTAGQNLTAAS 403

RESULT 9
 US-10-085-959-252

Sequence 252, Application US/10085959
 Publication No. US20030165870A1
 GENERAL INFORMATION:

APPLICANT: Blatner, Frederick R.
 APPLICANT: Welch, Rodney A.
 APPLICANT: Burland, Valerie D.
 TITLE OF INVENTION: No. US20030165870A1 Sequence of E. Coli CFT073
 FILE REFERENCE: 960296, 97648
 CURRENT APPLICATION NUMBER: US/10/085, 959
 CURRENT FILING DATE: 2002-03-01
 PRIOR APPLICATION NUMBER: 60/242, 412
 PRIOR FILING DATE: 2000-10-19
 NUMBER OF SEQ ID NOS: 255
 SOFTWARE: Patentin version 3.1
 SEQ ID NO 252
 LENGTH: 2834
 TYPE: PRT
 ORGANISM: Escherichia coli

FEATURE:
 NAME/KEY: misc feature
 LOCATION: (493)..
 OTHER INFORMATION: The 'Xaa' at location 493 stands for Asn, Asp, His, or Tyr.
 NAME/KEY: misc feature
 LOCATION: (1158)..
 OTHER INFORMATION: The 'Xaa' at location 1158 stands for Asp, or Tyr.
 NAME/KEY: misc feature
 LOCATION: (1159)..
 OTHER INFORMATION: The 'Xaa' at location 1159 stands for Leu.
 NAME/KEY: misc feature
 LOCATION: (1168)..
 OTHER INFORMATION: The 'Xaa' at location 1168 stands for Ser, or Asn.
 NAME/KEY: misc feature
 LOCATION: (1270)..
 OTHER INFORMATION: The 'Xaa' at location 1270 stands for Thr, or Pro.
 NAME/KEY: misc feature
 LOCATION: (13480)..
 OTHER INFORMATION: Unsure
 NAME/KEY: misc feature
 LOCATION: (31038)..
 OTHER INFORMATION: Unsure
 NAME/KEY: misc feature
 LOCATION: (31042)..
 OTHER INFORMATION: Unsure
 NAME/KEY: misc feature
 LOCATION: (31770)..
 OTHER INFORMATION: Unsure

OTHER INFORMATION: Unsure
 NAME/KEY: misc feature
 LOCATION: (31799)..
 OTHER INFORMATION: Unsure
 NAME/KEY: misc feature
 LOCATION: (44922)..
 OTHER INFORMATION: Unsure
 US-10-085-959-252

Query Match 10.7%; Score 82.5; DB 14; Length 2834;
 Best Local Similarity 27.8%; Pred. No. 45;
 Matches 42; Conservative 16; Mismatches 54; Indels 39; Gaps 7;

QY 8 AFAIVSGSALAGVVPQWGG---GNHGGN---SSGPDSTIYQGSANA--ALA 58
 DB 283 ASGRITVNNSLATGAVTAKGQVTLTGPHKAGNLSVSRDIVLS---NGTLNSDKDLS 339
 QY 59 LOSDARKYDQVTRVVTHEMAHAGAGADNSTIELTNGFRNNATIDQWNAKNSDITVGY 118
 DB 340 LTAGGRITQ-----NEKLTAGRDVTLAAKNITD-----TASQINAAADIVTV--- 383
 QY 119 GGNNAALVNOTASDSSVWVROVFGNNATAN 149
 DB 384 -----ASDITLTGQITTAGQNLTAAS 403

RESULT 10

US-09-996-194-16

Sequence 16, Application US/09996194
 Patent No. US20020151696A1
 GENERAL INFORMATION:

APPLICANT: Bandaru, Rajasekhkar
 TITLE OF INVENTION: 84242, 8035, 55304, 52999, and 21999,
 FILE REFERENCE: 35800/240590
 CURRENT APPLICATION NUMBER: US/09/996, 194
 CURRENT FILING DATE: 2001-11-28
 PRIOR APPLICATION NUMBER: 60/250, 348
 PRIOR FILING DATE: 2000-11-30
 PRIOR APPLICATION NUMBER: 60/250, 073
 PRIOR FILING DATE: 2000-11-30
 PRIOR APPLICATION NUMBER: 60/253, 878
 PRIOR FILING DATE: 2000-11-29
 PRIOR APPLICATION NUMBER: 60/250, 338
 PRIOR FILING DATE: 2000-11-30
 NUMBER OF SEQ ID NOS: 22
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 16
 LENGTH: 65
 TYPE: PRT
 ORGANISM: Artificial Sequence

FEATURE:
 OTHER INFORMATION: Pfam consensus sequence
 US-09-996-194-16

Query Match 10.6%; Score 82; DB 9; Length 65;
 Best Local Similarity 58.1%; Pred. No. 0.4;
 Matches 18; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

QY 51 GSANALALQSDARKYDQVTRVVTHEMAH 81
 DB 22 GVINIPANITSRNHYDQVTRVVTHEIAHA 52

RESULT 11

US-09-820-843A-21

Sequence 21, Application US/09820843A
 Publication No. US2003003963A1
 GENERAL INFORMATION:

APPLICANT: Council of Scientific and Industrial Research
 TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTE
 FILE REFERENCE: Q63915

```

1 CURRENT APPLICATION NUMBER: US/09/820,843A
2
3 CURRENT FILING DATE: 2001-03-30
4
5 NUMBER OF SEQ ID NOS: 118
6
7 SOFTWARE: PatentIn version 3.0
8
9 SEQ ID NO 21
10
11 LENGTH: 354
12
13 TYPE: PRT
14
15 ORGANISM: M. tuberculosis
16
17 FEATURE:
18
19 NAME/KEY: misc_feature
20
21 OTHER INFORMATION: PPE
22
23 NAME/KEY: misc_feature
24
25 OTHER INFORMATION: gi|1781260
26
27 US-09-820-843A-21

```

Query Match	10.6%	Score 82;	DB 10;	Length 354;
Best Local Similarity	25.2%	Pred. NO. 3.5;		
Matches 31; Conservative	11;	Mismatches 47;	Indels 34;	Gaps 4;

QY 27 GCGGNHNHGGGSSGPSTLSIYOYGANNAALALSDARKYDOLVTRVVTHEMALAGGAD 86
:::
Db GNNGNFNFPGSGNTG-SNNIGFGMTGSCN-----FGFGNTGNN 81
87 NSTIELTONGFRNRATTIDQMANKNSDIITYGYGANNALVNQTASDSSVMWROYGFNGNA 146
:::
Db 82 NIGTGLTGDS--QITGLLNSGSNGTIGPNSGTGNVLFLNSGTGN-----VGFGNSG 131
:::
QY 147 TAN 149
:::
Db 132 TAN 134

RESULT 12
US-10-369-493-20619

Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ. ID NOS: 47374
SEQ. ID NO 20619
LENGTH: 486
TYPE: PRT
ORGANISM: *Rhodospseudomonas palustris*
FEATURES:
NAME/KEY: unsure
LOCATION: (1)..(486)
OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-20619

	Query Match	10.4%	Score 80.5;	DB 15;	Length 486;
	Best Local Similarity	23.5%;	Pred. No.7.5;		
	Matches	35;	Conservative	22;	Mismatches 47; Indels 45; Gaps 7;
Oy	11 AIVVSGS-----ALAGVPQWGGGNNHNGSNSGP-----DSTLSIYGSANNAAL	57			
Dd	86 SVVVAGTDHYADIVAVGLPSLGGVKKAAAYIDSSGFPAPFMDLASDPFAPEFGAADDG-	144			
Oy	58 ALGDARKDYQLVTRVVTHMMAHG--QGADNSTIEILONG---FNNAATIQWAKNSKD	112			
Dd	145 -----VIHTPAVGGRRG-----LIQGNNLLAQSSLVDAMRRTLEAD	183			
Oy	113 ITVG-----OYGNNNALVNQTASDSSEVM	136			

Db 184 VNIGMLPLFHVTLGLMLTLQAGGASVI 212

```

RESULT 13
US-10-156-761-13039
; Sequence 13039, Application US/10156761.
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMDURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHITUKU
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 13039
; LENGTH: 562
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-13039

```

Query Match	10.3%	Score 79.5;	DB 14;	Length 562;
Best Local Similarity	23.5%;	Pred. No. 11;		
Matches 53;	Conservative 20;	Mismatches 56;	Indels 97;	Gaps 11;

```

OY 15 SGGALAGVFPWGGGNNHNG---GNS-----SGPDT--LSIYOYGAAN-----55
Db 143 SGSPSYASLPKGIAGSENAGLITGSSYIGGDDGTGTDSIGRLNLSYQAPANGSFGEN 202
OY 56 -----ALALOS-----DARKYDOLTVR-----VTHEMAH-----80
Db 203 IRIHFMRSDAKTMQALYIPVOTSNNKKGYDATRTBPSTGYSWKPVWQGHYEAANDHGS 262
OY 81 -----ACGG-----ADNSTIELTONGFRNNATTIDOMAKNS 111
Db 263 VHGWMELEVAATGALQGRLEIPFIIDOSKLSANAVDTTIGIAMTNIRTN--LADFSIRAQ 320
OY 112 DITVGYGGNNAL-----VNO-----TASDSSVMVROYGFGNN 145
Db 321 NITSGYAGONTPLARIGNNVTNNKULLSISDDMONSGRRKGFRRAN 366

```

```

1      RESULT 14
2      US-10-369-493-3641
3      ; Sequence 3641, Application US/10369493
4      ; Publication No. US2003023675A1
5      ; GENERAL INFORMATION:
6      ; APPLICANT: Cao, Yongwei
7      ; APPLICANT: Hinkle, Gregory J.
8      ; APPLICANT: Slater, Steven C.
9      ; APPLICANT: Goldman, Barry S.
10     ; APPLICANT: Chen, Xianning
11     ; TITLE OR INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
12     ; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
13     ; FILE REFERENCE: 38-10(52052)B
14     ; CURRENT APPLICATION NUMBER: US/10/369,493
15     ; CURRENT FILING DATE: 2003-02-28
16     ; PRIOR APPLICATION NUMBER: US 60/360,039
17     ; PRIOR FILING DATE: 2002-02-21
18     ; NUMBER OF SEQ ID NOS: 47374
19     ; SEQ ID NO 3641
20     ; LENGTH: 276
21     ; TYPE: PRT

```

ORGANISM: Neurospora crassa
US-10-369-493-3641

Query Match 10.2%; Score 79; DB 15; Length 276;
Best Local Similarity 23.5%; Pred. No. 5.2;
Matches 42; Conservative 18; Mismatches 69; Indels 50; Gaps 6;

QY 19 LAGVVPQWGGGNGNG-----CGNSGPDSTLSIYQGSANALALQS 61
DB 3 LAGRVAVLTGGGTGIGLMAIQAALVANGAKYTIAGRNEKLDKVEITNKDVEGETIALQA 62
QY 62 DAKRYDQ--LVTRVTHE-----MAAGGADNSTI-----ELTQNGFRN-NATIDQ 105
DB 63 DVTKEEDIALVKEIESREKCLILVNNAGISSSVTTTEASDPKELKHNLPDENENATFDD 122
QY 106 WNAKNS-----DITVQYGGANALVNOTASDSVYWRQVFGNNAT 147
DB 123 WTETRTNVASIYFMTSAFLPLQASTERHFGWSGTVVNISSISGVKSAQHPSYNAS 181

RESULT 15

US-10-032-585-7876
Sequence 7876, Application US/10032585
Publication No. US20030180953A1
GENERAL INFORMATION:
APPLICANT: Terry, Roemer D.
APPLICANT: Bo, Jiang
APPLICANT: Charles, Boone
APPLICANT: Howard, Bussey
TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
FILE REFERENCE: 10182-005-999
CURRENT APPLICATION NUMBER: US/10/032,585
CURRENT FILING DATE: 2001-12-20
NUMBER OF SEQ ID NOS: 8000
SOFTWARE: PatentIn version 3.1
SEQ ID NO 7876
LENGTH: 688
TYPE: PRT
ORGANISM: Candida albicans
US-10-032-585-7876

Query Match 10.2%; Score 79; DB 14; Length 688;
Best Local Similarity 22.7%; Pred. No. 17;
Matches 27; Conservative 24; Mismatches 60; Indels 8; Gaps 3;

QY 30 GNNHGGNSGPDSTLSIYQGSANALALQSDARKYDQVTRVYTHMAHAGGADNST 89
DB 488 GNNNGSGNSGTTNSNNYN---NKSISKNEIDDDDLNPTSITNNTGTNNNNKSP 543
QY 90 IE-LTQNGFRNATIDQWNAKNSDITVQYGGANALVNOTASDSVYWRQVFGNNAT 147
DB 544 AKSKKSNFQNSNSALNLDKSKLKINT--NEITNISSETSNSSSPVILNHHGRSS 599

Search completed: March 11, 2004, 19:18:38
Job time : 25.6 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:24:14 / Search time 10.3 Seconds
(without alignments)
1410.186 Million cell updates/sec

Title: US-09-543-407-20

Perfect score: 774
Sequence: 1 MLLKLVAAFAAIVSGSALA.....DSSVMVROYGFGNNATANQY 151

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	689	89.0	151	2 JC6039	fimbria protein ag
2	689	89.0	151	2 A10635	major curlin chain
3	528	68.2	151	2 S70788	curlin protein csg
4	506.5	65.4	152	2 D90806	curlin major subun
5	506.5	65.4	152	2 H85665	hypothetical prote
6	105.5	13.6	145	2 AD3143	conserved hypotet
7	105.5	13.6	145	2 H98144	hypothetical prote
8	99	12.8	573	2 C86266	F319.21 protein -
9	96.5	12.5	151	2 S70787	curlin nucleator p
10	96.5	12.5	151	2 C90806	minor curlin subun
11	96.5	12.5	151	2 G85665	curlin minor chain
12	96	12.4	2174	2 E95965	hypothetical glyci
13	92.5	12.0	1567	2 S11672	ice nucleation pro
14	91.5	11.8	1332	2 S07053	ice nucleation pro
15	91	11.8	602	1 P10221	leishmanolysin (EC
16	91	11.8	645	2 F70825	probable PPE prote
17	90	11.6	639	2 C42049	leishmanolysin (EC
18	89.5	11.6	374	2 T03875	probable homeobox
19	89.5	11.6	375	2 T03874	probable homeobox
20	89.5	11.6	447	2 G84687	probable disease r
21	89	11.5	652	2 E97857	cell surface anti
22	88.5	11.4	151	2 UC6040	fimbria protein ag
23	88.5	11.4	151	2 AH0635	nucleation compone
24	88.5	11.4	1588	2 A86036	probable adhesin Z
25	88.5	11.4	1588	2 H91388	probable adhesin E
26	88	11.4	590	1 A45621	leishmanolysin (EC
27	88	11.4	599	2 B42049	leishmanolysin (EC
28	88	11.4	599	2 A44951	leishmanolysin (EC
29	88	11.4	646	1 S19916	leishmanolysin (EC

30	87.5	11.3	575	2 S35327	protein kinase gsg
31	87	11.2	1748	2 S42136	cnb protein - tet
32	86	11.1	1635	2 A10452	hemolysin (importe
33	85.5	11.0	438	2 T35789	probable secreted
34	85	11.0	615	2 E70663	probable PPE prote
35	85	11.0	891	2 E96590	hypothetical prote
36	85	11.0	3705	2 AD0123	probable autocrine
37	84.5	10.9	188	1 Y0ECF2	F7-2 fimbrial prot
38	84.5	10.9	1034	2 JC2143	ice nucleation act
39	84.5	10.9	1258	2 J00188	ice nucleation pro
40	84	10.9	494	2 T15502	hypothetical prote
41	83	10.7	407	2 F87494	hypothetical prote
42	83	10.7	409	2 T20847	hypothetical prote
43	83	10.7	590	2 E70946	probable PPE prote
44	83	10.7	678	2 A70762	probable PPE prote
45	83	10.7	3190	2 T13828	CRAB-binding prote

ALIGNMENTS

RESULT 1

JC6039
fimbria protein agfa precursor - Salmonella enteritidis
C/Species: Salmonella enteritidis
C/Date: 31-Dec-1996 #sequence revision 31-Dec-1996 #text_change 08-Oct-1999
C/Accession: JC6039; PC6015; A44898
R/Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Baner, P.A.; Kay, W.W.
J. Bacteriol. 178, 662-667, 1996
A/Title: Salmonella enteritidis agfaAC operon encoding thin, aggregative fimbriae.
A/Reference number: JC6039; PMID:96146512; PMID:8550497
A/Accession: JC6039
A/Molecule type: DNA
A/Residues: 1-151 <COL>
A/Cross-references: GB:U43280; NID:G1184712; PID:AA43599.1; PID:G1184714
A/Accession: PC6015
A/Molecule type: protein
A/Residues: 21-52 <CO2>
A/Experimental source: strain 27655-3b
A/Note: the authors translated the codon ACG for residue 44 as Ile
R/Collinson, S.K.; Emdy, L.; Muller, K.H.; Trust, T.J.; Kay, W.W.
J. Bacteriol. 173, 4773-4781, 1991
A/Title: Purification and characterization of thin, aggregative fimbriae from Salmonella
A/Reference number: A44898; PMID:91310586; PMID:1677357
A/Accession: A44898
A/Contents: 27655
A/Status: preliminary
A/Molecule type: protein
A/Residues: 21-33 <CO3>
A/Note: sequence extracted from NCBI backbone (NCBIP:45936)
C/Genetics:
A/Gene: agfa
C/Function:
A/Description: major component of thin aggregative fimbriae
A/Note: fimbriae bind to fibronectin, plasminogen, tissue plasminogen activator
C/Keywords: fimbria
F/1-20/Domain: signal sequence #status predicted <SIG>
F/21-151/Product: fimbria protein agfa #status experimental <MAT>
Query Match 89.0%; Score 689; DB 2; Length 151;
Best Local Similarity 90.7%; Pred. No. 3.1e-51;
Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 1 MLLKLVAAFAAIVSGSALAGVPPMGGGGNNCGSSGPDSTLSIYOGSANAALALQ 60
DB 1 MLLKLVAAFAAIVSGSALAGVPPMGGGGNNCGSSGPDSTLSIYOGSANAALALQ 60
QY 61 SDARKYDQLYTRVVTHEMAHAGQADNSTIELLTONGFRNNATIDQNNASNDITVQYGG 120
DB 61 SDARKSETTITQSGNGAGDVGGADNSTIELLTONGFRNNATIDQNNASNDITVQYGG 120
QY 121 NNAALVNOTASDSSVMVROYGFGNNATANQY 151
DB 121 NNAALVNOTASDSSVMVROYGFGNNATANQY 151

Db 121 NNAALVNOTASDSSVMVROVGFNNATATNOY 151

RESULT 2

major curlin chain precursor (imported) - Salmonella enterica subsp. enterica serovar Typhimurium

A:Species: Salmonella enterica subsp. enterica serovar Typhimurium

A:Note: this species has also been called Salmonella typhimurium

C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002

C:Accession: A10635

R:Parthill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Cornerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Mole, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A:Reference number: AB0502; MUID:21534947; PMID:11677608

A:Accession: A10635

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-151 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD08268.1; PID:gl6502315; GSPDB:GN00176

C:Genetics: STY1181

Query Match 89.0%; Score 689; DB 2; Length 151;

Best Local Similarity 90.7%; Pred. No. 3.1e-51;

Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 1 MKLTKVAFPAIVSGSALAGVPPQMGCGNHNCGNSGPDSTLSIYOGSANAALALQ 60

Db 1 MKLTKVAFPAIVSGSALAGVPPQMGCGNHNCGNSGPDSTLSIYOGSANAALALQ 60

QY 61 SDARKYDQVTRVVTTHMAHAGGADNSTIELTONGFPNNATIDOMNAKNSDITVGOYGG 120

Db 61 SDARKSETTITQSGGNADVDGQGDNSTIELTONGFPNNATIDOMNAKNSDITVGOYGG 120

QY 121 NNAALVNOTASDSSVMVROVGFNNATATNOY 151

Db 121 NNAALVNOTASDSSVMVROVGFNNATATNOY 151

RESULT 3

curlin protein csgA precursor - Escherichia coli (strain K-12)

N:Alternate names: csgA protein; major curlin protein

C:Species: Escherichia coli

C:Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 01-Mar-2002

C:Accession: S70788; G64846; S31202; S34560; S34559

R:Hammar, M.; Arngvist, A.; Bian, Z.; Olsen, A.; Normark, S.

Mol. Microbiol. 18, 661-670, 1995

A:Title: Expression of two csg operons is required for production of fibronectin- and Cc

A:Reference number: S70783; MUID:9641468; PMID:8817489

A:Accession: S70788

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-151 <HAM>

A:Cross-references: EMBL:X90754; NID:g147558; PIDN:CAA62282.1; PID:g1147564

A:Experimental source: strain K12, substrain W3110

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co

A:Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: G64846

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-151 <BLAT>

A:Cross-references: GB:AE000205; GB:U00096; NID:g1787265; PIDN:AACT4126.1; PID:g1787279;

A:Experimental source: strain K-12, substrain MG1655

R:Olson, A.; Arngvist, A.; Hammar, M.; Sukupolvi, S.; Normark, S.

Mol. Microbiol. 7, 523-536, 1993

A:Title: The Rpos sigma factor relieves H-NS-mediated transcriptional repression of csg

A:Reference number: S31202; MUID:93211294; PMID:8459772

A:Accession: S31202

A:Molecule type: DNA

A:Residues: 1-6, 'V', 8-151 <OLSI>

A:Cross-references: EMBL:L04979

A:Accession: S34560

A:Molecule type: protein

A:Residues: 21-42; 44-50 <OLS2>

R:Olson, A.N.; Arngvist, A.M.

submitted to the EMBL Data Library, October 1992

A:Reference number: S34559

A:Accession: S34559

A:Molecule type: DNA

A:Residues: 1-133, 'RORDSGWLV' <OLS3>

A:Cross-references: EMBL:L04979; NID:g290424; PIDN:AAA23616.1; PID:g290425

A:Experimental source: strain K-12, substrain W3110

C:Genetics:

A:Gene: csgA

A:Map position: 23.15

C:Function:

A:Description: major component of wild-type curli; interaction between CsgA and CsgB tr

A:Note: curli are thin, coiled fibers expressed on the surface of Escherichia coli that

F:1-20/Domain: signal sequence #status predicted <SIG>

F:21-151/Product: curlin #status experimental <MAT>

Query Match 68.2%; Score 528; DB 2; Length 151;

Best Local Similarity 68.9%; Pred. No. 1.2e-37;

Matches 104; Conservative 20; Mismatches 27; Indels 0; Gaps 0;

QY 1 MKLTKVAFPAIVSGSALAGVPPQMGCGNHNCGNSGPDSTLSIYOGSANAALALQ 60

Db 1 MKLTKVAFPAIVSGSALAGVPPQMGCGNHNCGNSGPDSTLSIYOGSANAALALQ 60

QY 61 SDARKYDQVTRVVTTHMAHAGGADNSTIELTONGFPNNATIDOMNAKNSDITVGOYGG 120

Db 61 TDARNSDLITTHGGGNADVDGQSDSDITLQGFNSATLIDOMNAKNSDITVGOYGG 120

QY 121 NNAALVNOTASDSSVMVROVGFNNATATNOY 151

Db 121 GNGAAVDDTASNSVNTVQVGFNNATATNOY 151

RESULT 4

curlin major subunit CsgA (imported) - Escherichia coli (strain O157:H7, substrain RIMD

C:Species: Escherichia coli

C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001

C:Accession: D90806

R:Hayashi, T.; Makino, K.; Onishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G

gasekawa, N.; Yasunaga, T.; Kuwara, S.; Shiba, T.; Hatori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen

A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: D90806

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-152 <HAY>

A:Cross-references: GB:BA000007; PIDN:BA034843.1; PID:g13360880; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RIMD 0509952

C:Genetics:

A:Gene: ECs1420

Query Match 65.4%; Score 506.5; DB 2; Length 152;

Best Local Similarity 67.1%; Pred. No. 7.7e-36;

Matches 102; Conservative 21; Mismatches 28; Indels 1; Gaps 1;

QY 1 MKLTKVAFPAIVSGSALAGVPPQMGCGNHNCGNSGPDSTLSIYOGSANAALAL 59

Db 1 MKLTKVAFPAIVSGSALAGVPPQMGCGNHNCGNSGPDSTLSIYOGSANAALAL 60

QY 60 SDARKYDQVTRVVTTHMAHAGGADNSTIELTONGFPNNATIDOMNAKNSDITVGOYGG 119

Db 61 QADARNSDLTITQHGCGAGVQGGSDSDSLDTQGFSGSATLDDQNGKDSMTVKQFG 120
 QY 120 GNNALVNOTASDSSVWVROVGFGNATANOY 151
 Db 121 GNGAALVDTASNSTVNTQVGFGNATANOY 152

RESULT 5

H85665
 hypothetical protein csgA [imported] - Escherichia coli (strain O157:H7, substrain EDL93)

C/Species: Escherichia coli
 C/Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001

C/Accession: H85665
 R/Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew

Miller, L.; Grobleck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoudis, K.; Apodaca, Nature 409, 529-533, 2001

A/Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A/Reference number: AB5480; MUID:21074935; PMID:11206551

A/Accession: H85665

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-152 <STO>

A/Cross-references: GB:AB005174; NID:g12514574; PIDN:AA655788.1; GSPDB:GN00145; UMGD:216

A/Experimental source: strain O157:H7, substrain EDL933

C/Genetics:

A/Map position: csgA

Query Match 65.4%; Score 506.5; DB 2; Length 152;
 Best Local Similarity 67.1%; Pred. No. 7.7e-36;
 Matches 102; Conservative 21; Mismatches 28; Indels 1; Gaps 1;
 QY 1 MKLTKAFAAIYVSGSALAGVPPQW-GGGGNNHGGNSGPDSTLSIYQGSANALAL 59
 Db 1 MKLTKAFAAIYVSGSALAGVPPQW-GGGGNNHGGNSGPDSTLSIYQGSANALAL 60
 QY 60 QSDARKYQDLVTRVVTHEMAHAGGADNSTIELTQNGFNATITDQMAKNSDITVGYG 119
 Db 61 QADARNSDLTITQHGCGAGVQGGSDSDSLDTQGFSGSATLDDQNGKDSMTVKQFG 120
 QY 120 GNNALVNOTASDSSVWVROVGFGNATANOY 151
 Db 121 GNGAALVDTASNSTVNTQVGFGNATANOY 152

RESULT 6

ADJ143
 conserved hypothetical protein Atu4768 [imported] - Agrobacterium tumefaciens (strain C5

C/Species: Agrobacterium tumefaciens

C/Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002

C/Accession: ADJ143

R/Wood, D.W.; Secubal, J.C.; Kaul, R.; Monke, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L

erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavina, T.; Levy, R.; Li, M.; McClellan

; Karp, P.; Romero, P.; Zhang, S. Science 294, 2317-2323, 2001

A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,

ster, E.W.

A/Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A/Reference number: AB2577; MUID:21608550; PMID:11743193

A/Accession: ADJ143

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-145 <KUR>

A/Cross-references: GB:AB008689; PIDN:AA145562.1; PID:g17743277; GSPDB:GN00187

A/Experimental source: strain C58 (Dupont)

C/Genetics:

A/Map position: linear chromosome

Query Match 13.6%; Score 105.5; DB 2; Length 145;
 Best Local Similarity 23.2%; Pred. No. 0.049;
 Matches 36; Conservative 27; Mismatches 61; Indels 31; Gaps 4;

QY 3 LKVAFAAIYVSGSALAGVPPQW-GGGGNNHGGNSGPDSTLSIYQGSANA 55
 Db 1 MKRSFISALVALVGLSAAAPAMANDVRIEYQGSNSAGAOEGYGNIRIYQNGYNR 60
 QY 56 ALALQSDARKYQDLVTRVVTHEMAHAGGADNSTIELTQNGFNATITDQMAKNSDITV 115
 Db 61 IVGHQ-----YGR-----HNL SAVGQEGHDNYGSTTQNGRNVAGI----- 96

QY 116 GOYGNALVNOTASDSSVWVROVGFGNATANOY 150
 Db 97 GQFGSNHTITLTQDNGNIAAGVYGRGCSANVSO 131

RESULT 7

H98144
 hypothetical protein AGR_L_228 [imported] - Agrobacterium tumefaciens (strain C58, Cere

C/Species: Agrobacterium tumefaciens

C/Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002

C/Accession: H98144

R/Goodner, B.; Hinkle, G.; Galtung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman

A.; Liu, F.; Mollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.

Science 294, 2323-2328, 2001

A/Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum

A/Reference number: A97359; MUID:21608551; PMID:11743194

A/Accession: H98144

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-145 <KUR>

A/Cross-references: GB:AB007870; PIDN:AAK86882.1; PID:g15158413; GSPDB:GN00170

C/Genetics:

A/Map position: linear chromosome

Query Match 13.6%; Score 105.5; DB 2; Length 145;
 Best Local Similarity 23.2%; Pred. No. 0.049;
 Matches 36; Conservative 27; Mismatches 61; Indels 31; Gaps 4;
 QY 3 LKVAFAAIYVSGSALAGVPPQW-GGGGNNHGGNSGPDSTLSIYQGSANA 55
 Db 1 MKRSFISALVALVGLSAAAPAMANDVRIEYQGSNSAGAOEGYGNIRIYQNGYNR 60
 QY 56 ALALQSDARKYQDLVTRVVTHEMAHAGGADNSTIELTQNGFNATITDQMAKNSDITV 115
 Db 61 IVGHQ-----YGR-----HNL SAVGQEGHDNYGSTTQNGRNVAGI----- 96
 QY 116 GOYGNALVNOTASDSSVWVROVGFGNATANOY 150
 Db 97 GQFGSNHTITLTQDNGNIAAGVYGRGCSANVSO 131

RESULT 8

C86266
 F3819.21 protein - Arabidopsis thaliana

C/Species: Arabidopsis thaliana (mouse-ear cress)

C/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Nov-2001

C/Accession: C86266

R/Theologis, A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.

ansen, N.F.; Hughes, B.; Huizar, L. Nature 408, 816-820, 2000

A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lures, J.S.; Maiti, R.; Marziani

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Xu, G.; Frazer, C.M.; Venter, J.C.; Davis, R.W.

A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A/Reference number: AB6141; MUID:21016719; PMID:11130712

A/Accession: C86266

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-573 <STO>

A/Cross-references: GB:AB005172; NID:g4850402; PIDN:AAJ31072.1; GSPDB:GN00141

C/Genetics:

A:Map position: 1

Query Match 12.8%; Score 99; DB 2; Length 573;
 Best Local Similarity 25.4%; Pred. No. 0.81;
 Matches 34; Conservative 24; Mismatches 46; Indels 30; Gaps 6;

QY 13 VVSGSALAGVVPQMGCGNHNCGNSGPDSTLSIYOGSANAALALQSDARKYDQVTR 72
 DB 84 IVSGGTVEG---KYNDGCHNG---ISGPDTRSDVYPPQASSFGAKGLNID----- 127

QY 73 VVTHEMAHAGAGADNSTELTQNGFRNNA-TIDQNNANUS-----DITVGYGGANN 122
 DB 128 IOSNKLIAOOG---STTVLNNHGFSGNAVNVPEMPVHNSVCGAPQGAQQLPVSGMSVNP 183

QY 123 AALVNOTASDSSVM 136
 DB 184 NVMMNKSPTQSFVV 197

RESULT 9

curlin nucleator protein csGB precursor - Escherichia coli (strain K-12)
 S70787
 N:Alternate names: csGB protein; curlin nucleation component; minor curlin protein
 C:Species: Escherichia coli
 C>Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 01-Mar-2002
 C/Accession: S70787; F64846
 R:Hammar, M.; Arngvist, A.; Bian, Z.; Olsen, A.; Normark, S.
 Mol. Microbiol. 18, 661-670, 1995
 A>Title: Expression of two csG operons is required for production of fibronectin- and Cc
 A/Reference number: S70783; MUID:96414468; PMID:8817489
 A/Accession: S70787
 A>Status: nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-151 <HAM>

A/Cross-references: EMBL:X90754; NID:G1147558; PIDN:CAA62281.1; PID:G1147563
 A/Experimental source: strain K12, substrain W3110
 A/Note: The nucleotide sequence was submitted to the EMBL Data Library, August 1995
 R:Battner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
 A.; Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A>Title: The complete genome sequence of Escherichia coli K-12.
 A/Reference number: A64720; MUID:9742617; PMID:9278503
 A/Accession: F64846
 A>Status: nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-151 <BLAT>

A/Cross-references: GB:AE00205; GB:U00096; NID:G1787265; PIDN:AACT4125.1; PID:G1787278;
 A/Experimental source: strain K-12, substrain MG1655
 C/Genetics:
 A:Gene: csGB
 A:Map position: 23.15
 C/Function:
 A>Description: minor component of wild-type curli; interaction between CsgA and CsgB tri
 A/Note: curli are thin, coiled fibers expressed on the surface of Escherichia coli that
 F:1-21/Dominant signal sequence #status predicted <SIG>
 F:22-151/Product: minor curlin chain #status predicted <MAT>

Query Match 12.5%; Score 96.5; DB 2; Length 151;
 Best Local Similarity 28.1%; Pred. No. 0.29;
 Matches 32; Conservative 15; Mismatches 62; Indels 5; Gaps 2;

QY 38 SSGPDSTLSIYOGSANAALALQSDARKYDQVTRVTHEMAHAGAGADNSTELTQNGF 97
 DB 21 AAGYDLANSEYF---AVNELSKSFNOAIIIGAGTNNSAQLRQGSKLAVVAQEGS 76

QY 98 RNNATIDQNNAKNSDITVGYGGNNAALVNOTASDSSVMVRYQFGNNATANQY 151
 DB 77 SNRAKIDQTDGYNL-AVIDQAGSANDASISQAGYGTAMTIQKSGNKANITQY 129

RESULT 10

C90806

minor curlin subunit precursor CsgB [imported] - Escherichia coli (strain O157:H7, subsp
 C:Species: Escherichia coli
 C/Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
 C/Accession: C90806
 R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G
 gaswara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 8, 11-22, 2001
 A>Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen
 A/Reference number: A99629; MUID:21156231; PMID:11258796
 A/Accession: C90806
 A>Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-151 <HAY>

A/Cross-references: GB:BA00007; PIDN:BA834842.1; PID:G13360879; GSPDB:GN00154
 A/Experimental source: strain O157:H7, substrain RIMD 0509952
 C/Genetics:
 A:Gene: EC81419

Query Match 12.5%; Score 96.5; DB 2; Length 151;
 Best Local Similarity 28.1%; Pred. No. 0.29;
 Matches 32; Conservative 15; Mismatches 62; Indels 5; Gaps 2;

QY 38 SSGPDSTLSIYOGSANAALALQSDARKYDQVTRVTHEMAHAGAGADNSTELTQNGF 97
 DB 21 AAGYDLANSEYF---AVNELSKSFNOAIIIGAGTNNSAQLRQGSKLAVVAQEGS 76

QY 98 RNNATIDQNNAKNSDITVGYGGNNAALVNOTASDSSVMVRYQFGNNATANQY 151
 DB 77 SNRAKIDQTDGYNL-AVIDQAGSANDASISQAGYGTAMTIQKSGNKANITQY 129

RESULT 11

curlin minor chain precursor, CsgA homolog [imported] - Escherichia coli (strain O157:H
 C:Species: Escherichia coli
 C/Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 C/Accession: G85665
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glaesner, J.D.; Rose, D.J.; Mayne
 filler, L.; Grobeck, E.J.; Davis, N.W.; Lm, A.; Dimalanta, E.; Petcamouls, K.; Apodaca,
 Nature 409, 529-533, 2001
 A>Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A/Reference number: A85480; MUID:21074935; PMID:11206551
 A/Accession: G85665
 A>Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-151 <STO>

A/Cross-references: GB:AE005174; NID:G12514573; PIDN:AA655787.1; GSPDB:GN00145; UWGP:21
 A/Experimental source: strain O157:H7, substrain EDL933
 C/Genetics:
 A:Gene: csGB

Query Match 12.5%; Score 96.5; DB 2; Length 151;
 Best Local Similarity 28.1%; Pred. No. 0.29;
 Matches 32; Conservative 15; Mismatches 62; Indels 5; Gaps 2;

QY 38 SSGPDSTLSIYOGSANAALALQSDARKYDQVTRVTHEMAHAGAGADNSTELTQNGF 97
 DB 21 AAGYDLANSEYF---AVNELSKSFNOAIIIGAGTNNSAQLRQGSKLAVVAQEGS 76

QY 98 RNNATIDQNNAKNSDITVGYGGNNAALVNOTASDSSVMVRYQFGNNATANQY 151
 DB 77 SNRAKIDQTDGYNL-AVIDQAGSANDASISQAGYGTAMTIQKSGNKANITQY 129

RESULT 12

E95965
 hypothetical glycine-rich protein [imported] - Sinorhizobium meliloti (strain 1021) mag
 C:Species: Sinorhizobium meliloti
 C/Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
 C/Accession: E95965
 R:Finan, T.M.; Weidner, S.; Wong, K.; Buhmester, J.; Chain, P.; Votholter, F.J.; Hernar
 Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
 A>Title: The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing end

A:Reference number: A95842; MUID:21396508; PMID:11481431
 A:Accession: E93965
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-2174 <KUR>
 A:Cross-references: GB:AL51985; PIDN:CA49389.1; PID:g15110875; GSPDR:GN00167
 A:Experimental source: strain 1021, megaplasmid pSymB
 R:Gallberg, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler, P.; Chaitin, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.; L.; Hyman, R.W.; Jones, T.
 Science 293, 668-672, 2001
 A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure, hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
 A:Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.
 A:Reference number: A96039; MUID:21368234; PMID:11474104
 A:Accession: annotation
 C:Genetic: annotation
 A:Gene: Smb21548
 A:Genome: plasmid

Query Match 12.4%; Score 96; DB 2; Length 2174;
 Best Local Similarity 27.0%; Pred. No. 6.5;
 Matches 40; Conservative 18; Mismatches 54; Indels 36; Gaps 7;

QY 11 AIVSSSALAGVPPQ--WGGGNHNGGNSGPDSTLSIYQGS-----ANAA----- 56
 DB 693 AATATAGAGVGILOSIIGGGG--GNATGGDAFGSGFOIGGGGGGYANTAVGFR 749
 QY 57 -TALOSDARKYDQLVTRVVTHEMAHAGGADNSTIELTQNGFRNNAITDQWAKNSDITV 115
 DB 750 GLTLTTQGHAGIAGVAVS-----GGGGTGTGTAAGTGTAS-----VAV 793

RESULT 13

S11672
 ice nucleation protein - *Xanthomonas campestris*
 C:Species: *Xanthomonas campestris*
 C:Date: 21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 26-Aug-1999
 C:Accession: S11672
 R:Zhao, J.; Orser, C.S.
 Mol. Gen. Genet. 223, 163-166, 1990
 A:Title: Conserved repetition in the ice nucleation gene *inx* from *Xanthomonas campestris*
 A:Reference number: S11672; MUID:91080859; PMID:2259339
 A:Accession: S11672
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1567 <ZHA>
 A:Cross-references: EMBL:X52970; NID:948531; PIDN:CA37140.1; PID:948532
 C:Superfamily: ice nucleation protein

Query Match 12.0%; Score 92.5; DB 2; Length 1567;
 Best Local Similarity 25.8%; Pred. No. 8.9;
 Matches 46; Conservative 27; Mismatches 56; Indels 49; Gaps 11;

QY 14 VGGSLALG-----VVPQMG--GNHN-----GGSSGPDSTLSIYQGSANALAL 59
 DB 205 VGGSTLTGADQSRVLVAGYSTETAGHSDLIAGYGTGAGSDSI-LAGYGTGTAAAR 263
 QY 60 QSDARKYDQLVLT-----RVVTHEMAHAGGADNSTIELTQNGFRNNAIT----- 103
 DB 264 STLTAGYGTGTQNGESRLTSGGSRATGSDSAV-----SGYGTGTGSSSLTAGY 319
 QY 104 -DQWAKNSDITVGYG-----GNNAALV-----NOTASDSSVMPQVGFNNATNQ 150
 DB 320 STGTAKGSDITAG-VGSTGTAGSDALTAGYGTGTAGSSESLT--AGYGTGTARK 374

RESULT 14

S07053

ice nucleation protein *inaA* - *Erwinia ananas*
 C:Species: *Erwinia ananas*
 C:Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 26-Aug-1999
 C:Accession: S07053
 R:Debe, K.; Matabe, S.; Emori, Y.; Watanabe, M.; Arai, S.
 FEBS Lett. 258, 297-300, 1989
 A:Title: An ice nucleation active gene of *Erwinia ananas*. Sequence similarity to those
 A:Reference number: S07053; MUID:90092494; PMID:2599095
 A:Accession: S07053
 A:Status: not compared with conceptual translation
 A:Molecule type: DNA
 A:Residues: 1-1322 <AB>
 A:Cross-references: GB:X17316; NID:9296095; PIDN:CA35194.1; PID:9296096
 C:Superfamily: ice nucleation protein

Query Match 11.8%; Score 91.5; DB 2; Length 1322;
 Best Local Similarity 29.1%; Pred. No. 8.9;
 Matches 37; Conservative 19; Mismatches 22; Indels 49; Gaps 9;

QY 34 GGGSSGPDSTLSIYQGSANALALQSDARKYDQLVTRVVTHEMAHAGGADNSTIELT 93
 DB 933 GSTTAGPDSST-IGYGTGTQTA-----GYSNLT-----AGYGS-----T 967
 QY 94 QNGFRNNAITDQWAKNSDITVGYG-----GNNAALV-----NOTASDSSVMPQVGF 143
 DB 968 QTG-----QENSDLTG-YGSTGTAGYSSSLIAGYGTGTAFKSTLM--AGYG 1013
 QY 144 NNATNQ 150
 DB 1014 SSQTARE 1020

RESULT 15

PL0221

Leishmanolysin (EC 3.4.24.36) precursor [validated] - *Leishmania major*
 N:Alternate names: promastigote surface proteinase; surface endopeptidase glycoprotein y
 C:Species: *Leishmania major*
 C:Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 15-Sep-2000
 C:Accession: PL0221; A27598; A60648
 J:Button, L.L.; McMaster, W.R.
 J. Exp. Med. 171, 589, 1990
 A:Reference number: PL0221
 A:Contents: extratum
 A:Accession: PL0221
 A:Molecule type: DNA
 A:Residues: 1-602 <BU>
 A:Cross-references: GB:Y00647; NID:99554; PIDN:CA46873.1; PID:99555
 A:Note: this is a revision to the sequence from reference A27598
 R:Button, L.L.; McMaster, W.R.
 J. Exp. Med. 167, 724-729, 1988
 A:Title: Molecular cloning of the major surface antigen of *Leishmania*.
 A:Reference number: A27598; MUID:88154764; PMID:3346625
 A:Accession: A27598
 A:Status: significant sequence differences
 A:Molecule type: DNA
 R:Bouvier, J.; Bordier, C.; Vogel, H.; Reichelt, R.; Etges, R.
 Mol. Biochem. Parasitol. 37, 235-246, 1989
 A:Title: Characterization of the promastigote surface protease of *Leishmania* as a membra
 A:Reference number: A60648; MUID:90114330; PMID:2608099
 A:Accession: A60648
 A:Molecule type: protein
 A:Residues: 101,'E',103-118,'SV',121-123 <BOU>
 A:Experimental source: strain LEW513
 R:Schlagenhauf, E.; Etges, R.; Metcalf, P.
 submitted to the Brookhaven Protein Data Bank, March 1997
 A:Reference number: A68135; PDB:1LMU
 A:Contents: annotation; X-ray crystallography, 1.86 angstroms, residues 100-407, 412-498,
 A:Note: strain LRC-L119
 C:Complex: homodimer
 C:Function:
 A:Description: catalyzes the hydrolysis of peptide bonds between two hydrophobic residue
 A:Note: the activated form can activate the proenzyme form
 C:Superfamily: Leishmanolysin

C:Keywords: blocked carboxyl end; cell adhesion; glycoprotein; homodimer; hydrolase; lig
 F;1-39/Domain: signal sequence #status predicted <SIG>
 F;40-100/Domain: activation peptide #status predicted <ATP>
 F;101-577/Product: leishmanolysin #status experimental <MAT>
 F;578-602/Domain: carboxyl-terminal propeptide #status predicted <CTP>
 F;48,264,268,334/Binding site: zinc, catalytic (Cys, His, His, His) (inhibited) #status
 F;100-101/Cleavage site: Val-Val (autolytic) #status experimental
 F;125-142,191-230,314-386,393-455,406-425,415-489,466-510,515-565,535-558/Disulfide bond
 F;264,268,334/Binding site: zinc, catalytic (His) (active) #status experimental
 F;265/Active site: Glu #status predicted
 F;300,407/Binding site: carboxylate (Asn) (covalent) #status experimental
 F;577/Modified site: GPI-anchor ethanolamine amidated carboxyl end (Asn) (in mature form)

Query Match 11.8; Score 91; DB 1; Length 602;
 Best Local Similarity 37.58; Pred. No. 4;
 Matches 27; Conservative 6; Mismatches 15; Indels 24; Gaps 3;
 QY 63 ARKYDQLVTRVVTHEMAHA---GQADNSTIELTQNGFRNNATIDQNAKNSDITV---115
 Db 251 ASRYDQLVTRVVTHEMAHALGFSGPFEDARIV-----ANVPVRGKNFDPVPIVS 301
 QY 116 -----GQYG 119
 Db 302 STAVAKAREQYG 313

Search completed: March 11, 2004, 18:42:10
 Job time : 11.3 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Comugen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:14:48 ; Search time 6.3 Seconds

(without alignments)
1248.031 Million cell updates/sec

Title: US-09-543-407-20

Perfect score: 774
Sequence: 1 MRLKVAFAIIVSGSALA.....DSSVMVQVFGNNATANY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	68.9	89.0	151	1 CSGB_SALTY	P55325 salmonella
2	52.8	68.2	151	1 CSGB_ECOLI	P28307 escherichia
3	506.5	65.4	152	1 CSGB_ECO57	Q93124 escherichia
4	96.5	12.5	151	1 CSGB_ECOLI	P39828 escherichia
5	92.5	12.0	1567	1 ICEN_XANCT	P18127 xanthomonas
6	91.5	11.8	1322	1 ICEN_PANAN	P20469 pantoea ana
7	91	11.8	602	1 GP63_LEIMA	P08148 leishmania
8	88.5	11.4	151	1 CSGB_SALTY	Q827M3 salmonella
9	88.5	11.4	151	1 CSGB_SALTY	P55326 salmonella
10	88	11.4	590	1 GP63_LEIDO	P23323 leishmania
11	88	11.4	590	1 GP63_LEIDO	P15706 leishmania
12	88	11.4	366	1 GP63_LEIME	P43150 leishmania
13	86.5	11.2	369	1 PS73_MYCAV	Q9K889 mycobacteri
14	84.5	10.9	188	1 FMF2_ECOLI	P02972 escherichia
15	84.5	10.9	1034	1 ICEN_PANAN	Q47879 pantoea ana
16	84.5	10.9	1258	1 ICEN_ERWHE	P16339 erwinaa her
17	83	10.7	678	1 YF48_MYCTU	Q10778 mycobacteri
18	82.5	10.7	592	1 CEA_CITFR	P04480 citrobacter
19	82.5	10.7	641	1 IMD_ARTGO	Q44052 arthrobacte
20	82.5	10.7	1028	1 OVO_DROME	P51521 drosophila
21	81.5	10.5	392	1 HME1_HUMAN	Q05525 homo sapien
22	81.5	10.5	1317	1 N145_YEAST	P49687 saccharomyc
23	81	10.5	1656	1 OMPB_RICJA	Q06653 r outer mem
24	80.5	10.4	1210	1 ICEN_PSEPL	P09815 pseudomonas
25	78.5	10.1	472	1 Y442_MYCTU	P21131 bacillus am
26	78.5	10.1	487	1 Y442_MYCTU	P42611 mycobacteri
27	78	10.1	147	1 HFAA_CAUCR	P27342 caulobacter
28	77.5	10.0	363	1 PER_COPCI	P28314 coprinus ci
29	77.5	10.0	254	1 PRIO_MESAU	P04273 mesocricetu
30	77.5	10.0	364	1 PER_ARTRA	P28313 athryomyces
31	77.5	10.0	499	1 YPZ5_METTF	P29579 methanobact
32	77	9.9	370	1 PS73_MYCTU	Q06643 mycobacteri
33	76	9.8	252	1 PRIO_RABIT	Q95211 oryctolagus

34	76	9.8	504	1 FLIC_SALMC	Q06381 salmonella
35	76	9.8	955	1 FRU_DROME	Q81n81 drosophila
36	76	9.8	1115	1 TBC2_CHLRE	Q8vxp3 chlamydomon
37	75.5	9.8	254	1 PRIO_CRIGR	Q60506 cricetus
38	75.5	9.8	576	1 CEAY_ECOLI	Q47112 escherichia
39	75.5	9.8	1196	1 ICEN_PSEBX	Q33479 pseudomonas
40	75	9.7	172	1 CH18_DROME	P07184 drosophila
41	75	9.7	389	1 ACUC_STYAM	Q99tc9 staphylococ
42	74.5	9.6	163	1 HCV_NATPH	P39442 natronomona
43	74.5	9.6	370	1 PGLI_PENOL	Q9y834 penicillium
44	74.5	9.6	663	1 DUS8_MOUSE	Q09112 mus musculu
45	74.5	9.6	760	1 YBIL_ECOLI	P75780 escherichia

ALIGNMENTS

RESULT 1
ID CSGB_SALTY STANDARD; PRT; 151 AA.
AC P55325;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Major curlin subunit precursor (Fimbriae SEF17).
GN CSGB OR AGFA OR STM1144 OR STY1181 OR T1776.
OS Salmonella typhimurium,
OS Salmonella typhi, and
OS Salmonella enteritidis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602, 601, 592;
RN (1)
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=SR-11;
RX MEDLINE=98117058; PubMed=9457880;
RA Rowling U., Bian Z., Hammar M., Sierrala W.D., Normark S.;
RT "Curli fibers are highly conserved between Salmonella typhimurium and
EScherichia coli with respect to operon structure and regulation.";
RL J. Bacteriol. 180:722-731(1998).
RN (2)
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
LT2.";
RL Nature 413:852-856(2001).
RN (3)
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Partholl J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Felwell T., Hamlin N., Haque A., Hien T.T., Holtroyd S., Jagels K.,
RA Krogg A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrett B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
RN (4)
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=TY2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;
RA Deng W., Liu S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodoyanni V., Schwartz D.C., Blattner F.R.;

```

RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RT and CT18."
RL J. Bacteriol. 185:2330-2337(2003).
RN
RP SEQUENCE FROM N.A.
RC SPECIES=S. enteritidis; STRAIN=27655-3B;
RX MEDLINE=96146512; PubMed=8550497;
RA Collinson S.K., Clouthier S.C., Doran J.L., Baner P.A., Kay W.W.;
RT "Salmonella enteritidis agfBAC operon encoding thin, aggregative
RT fimbriae."
RL J. Bacteriol. 178:662-667(1996).
RN
RP SEQUENCE OF 21-151 FROM N.A.
RC SPECIES=S. enteritidis; STRAIN=27655-3B;
RX MEDLINE=94013373; PubMed=8104955;
RA Doran J.L., Collinson S.K., Burtian J., Santos G., Todd E.C.D.,
RA Munro C.K., Kay C.M., Baner P.A., Peterkin P.I., Kay W.W.;
RT "DNA-based diagnostic tests for Salmonella species targeting agfA,
RT the structural gene for thin, aggregative fimbriae."
RL J. Clin. Microbiol. 31:2263-2273(1993).
RN
RP SEQUENCE OF 21-33.
RC SPECIES=S. enteritidis; STRAIN=27655-3B;
RX MEDLINE=91310586; PubMed=1677357;
RA Collinson S.K., Emeody L., Mueller K.-M., Trust T.J., Kay W.W.;
RT "Purification and characterization of thin, aggregative fimbriae from
RT Salmonella enteritidis."
RL J. Bacteriol. 173:4773-4781(1991).
CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN.
CC
CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AJ002301; CAA05317.1; -
DR EMBL; AF008749; AAL20074.1; -
DR EMBL; AL627269; CAD08268.1; -
DR EMBL; AE016840; AAC09399.1; -
DR EMBL; U43280; AAC43599.1; -
DR PIR; JC6039; JC6039.
DR StGene; SGI0608; csGA.
KM Fimbria; Signal; Complete proteome.
FT SIGNAL 1 20
FT CHAIN 21 151 MAJOR CURLIN SUBUNIT.
FT CONFLICT 134 151 SWMRQVGFNNATANY -> DSYTQVAS (IN
FT REF. 6).
SQ SEQUENCE 151 AA; 15305 MW; B7DAC0D16B21359 CRC64;

```

```

Query Match 89.0%; Score 689; DB 1; Length 151;
Best Local Similarity 90.7%; Pred. No. 3.9e-53;
Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

```

```

QY 1 MKLLKVAFAAIVVSGSALAGVVPOMGGGNNHNGSGSPDSTLSIYQGSANAAALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPOMGGGNNHNGSGSPDSTLSIYQGSANAAALQ 60
QY 61 SDARKYDQLVTRVTHMAHAGGADNSTIELTNGFRNNATTIDQNAKNSDITVGQYG 120
DB 61 SDARKSETITQSGYGADVGQADNSTIELTNGFRNNATTIDQNAKNSDITVGQYG 120
QY 121 NNAAALVNOTASDSSVMVRQVGFNNATANY 151
DB 121 NNAAALVNOTASDSSVMVRQVGFNNATANY 151

```

```

RESULT 2
CSGA_ECOLI
ID CSGA_ECOLI STANDARD; PRT; 151 AA.
AC P28307;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Major curlin subunit precursor.
GN CSGA OR B1042.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / W3110;
RX MEDLINE=93211294; PubMed=8459772;
RA Olsen A., Arngvist A.;
RT "The RpoS sigma factor relieves H-NS-mediated transcriptional
RT repression of csGA, the subunit gene of fimbriectin-binding curli in
RT Escherichia coli."
RL Mol. Microbiol. 7:523-536(1993).
RN
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MC4100;
RX MEDLINE=96414468; PubMed=8817489;
RA Hammar M., Arngvist A., Bian Z., Olsen A., Normark S.;
RT "Expression of two csG operons is required for production of
RT fimbriectin- and Congo red-binding curli polymers in Escherichia coli
RT K-12."
RL Mol. Microbiol. 18:661-670(1995).
RN
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97446617; PubMed=9278503;
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Sha Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Vano M., Horinouchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map."
RL DNA Res. 3:137-155(1996).
RN
RP SEQUENCE OF 21-40.
RC STRAIN=K12 / YMEI;
RX MEDLINE=93023873; PubMed=1357528;
RA Arngvist A., Olsen A., Pfeiffer J., Russell D.G., Normark S.;
RT "The Crl protein activates cryptic genes for curli formation and
RT fibronectin binding in Escherichia coli HB101."
RL Mol. Microbiol. 6:2443-2452(1992).
RN
RP SEQUENCE OF 21-31.
RC MEDLINE=91310586; PubMed=1677357;
RA Collinson S.K., Emeody L., Trust T.J., Kay W.W.;
RT "Purification and characterization of thin, aggregative fimbriae from
RT Salmonella enteritidis."
RL J. Bacteriol. 173:4773-4781(1991).
CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN.

```

CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; L04979; AAA23616.1; -
 DR EMBL; X80754; CAA62282.1; -
 DR EMBL; AE000205; AAC74126.1; -
 DR EMBL; D90741; BAA35832.1; -
 DR EMBL; D90742; BAA35840.1; -
 DR PIR; S70788; S70788.
 DR EcoGene; EG11489; CSGA.
 KW Fimbria; Signal; Complete proteome.
 FT SIGNAL 1 20
 FT CHAIN 21 151 MAJOR CURLIN SUBUNIT.
 FT CONFICT 7 7 A -> E (IN REF. 1).
 SQ SEQUENCE 151 AA; 15049 MW; C003470D208D395F CRC64;
 Query Match 68.2%; Score 528; DB 1; Length 151;
 Best Local Similarity 68.3%; Pred. No. 3.6e-39;
 Matches 104; Conservative 20; Mismatches 27; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAFAIVSGSALAGVVPWGNGGNGGNGSGPDSITLTIYOGSANAALALQ 60
 DB 1 MKLLKVAFAFAIVSGSALAGVVPWGNGGNGGNGSGPDSITLTIYOGGNSALALQ 60
 QY 61 SDARKYDQLVTRVTVTHMAHAGGADNSTIELTONGFRNNATIDQNNAKNSDITVQYGG 120
 DB 61 TDARNSDLITTHQGGNGADVGQGSDDSIDLTQGRGNSATIDQNNKNSDITVQYGG 120
 QY 121 NNAALVNOTASDSSVMTRQVGFNNATANQY 151
 DB 121 GNGAAVDQTASNSVTVTVQVFGNNATAHQY 151
 RESULT 3
 CSGA_ECO57 STANDARD; PRT; 152 AA.
 ID CSGA_ECO57 STANDARD; PRT; 152 AA.
 AC Q93024;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Major curlin subunit precursor.
 GN CSGA OR Z1676 OR ECS1420.
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 NC NCB1_TaxID=83334;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / ATCC 43895;
 RX MEDLINE=21218556; PubMed=11319125;
 RA Uhlrich G.A., Keen J.B., Elder R.O.;
 RT "Mutations in the csgD promoter associated with variations in curli
 RT expressions in certain strains of Escherichia coli O157:H7";
 RL Appl. Environ. Microbiol. 67:2367-2370(2001).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11286551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grothbeck E.J., Davis N.W., Iam A., Dimlantia E.T., Potamouis K.,
 RA Apodoca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
 RT Nature 409:529-533(2001).

RN (3)
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RIMD 0509952;
 RX MEDLINE=21156211; PubMed=11258796;
 RA Hayashi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohemorrhagic Escherichia coli
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22(2001)
 CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 CC FIBRONECTIN.
 CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AE275733; AAK53212.1; -
 DR EMBL; AE005315; AAG55788.1; -
 DR EMBL; AP002554; BAB34843.1; -
 DR PIR; D90806; D90806.
 DR PIR; H85665; H85665.
 KW Fimbria; Signal; Complete proteome.
 FT SIGNAL 1 20
 FT CHAIN 21 152 MAJOR CURLIN SUBUNIT.
 SQ SEQUENCE 152 AA; 15099 MW; EE2D2D94DDE91243 CRC64;
 Query Match 65.4%; Score 506.5; DB 1; Length 152;
 Best Local Similarity 67.1%; Pred. No. 2.6e-37;
 Matches 102; Conservative 21; Mismatches 28; Indels 1; Gaps 1;
 QY 1 MKLLKVAFAFAIVSGSALAGVVPWGNGGNGGNGSGPDSITLTIYOGSANAALALQ 59
 DB 1 MKLLKVAFAFAIVSGSALAGVVPWGNGGNGGNGSGPDSITLTIYOGGNSALALQ 60
 QY 60 QSDARKYDQLVTRVTVTHMAHAGGADNSTIELTONGFRNNATIDQNNAKNSDITVQYGG 119
 DB 61 QDARNSDLITTHQGGNGADVGQGSDDSIDLTQGRGNSATIDQNNKNSDITVQYGG 120
 QY 120 GNNALVNOTASDSSVMTRQVGFNNATANQY 151
 DB 121 GNGAAVDQTASNSVTVTVQVFGNNATAHQY 152
 RESULT 4
 CSGB_ECOLI STANDARD; PRT; 151 AA.
 ID CSGB_ECOLI STANDARD; PRT; 151 AA.
 AC P39828;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Minor curlin subunit precursor.
 GN CSGB OR B1041 OR Z1675 OR ECS1419.
 OS Escherichia coli, and
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 NC NCB1_TaxID=562; 83334;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=KL2 / MC4100;
 RX MEDLINE=96414468; PubMed=8817489;
 RA Hammar M., Argyriet A., Bian Z., Olsen A., Normark S.;
 RT "Expression of two csg operons is required for production of
 RT fibronectin- and Congo red-binding curli polymers in Escherichia coli

RT K-12." ;
 RL Mol. Microbiol. 18:661-670(1995).
 RL [2]
 RC SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 MAU B., Shao Y.;
 RT "The complete genome sequence of *Escherichia coli* K-12." ;
 RL Science 277:1453-1474(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RP STRAIN=K12;
 RX MEDLINE=97061202; PubMed=8905232;
 RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
 RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horiiuchi T.;
 RT "A 718-kb DNA sequence of the *Escherichia coli* K-12 genome
 corresponding to the 12.7-28.0 min region on the linkage map." ;
 RL DNA Res. 3:137-155(1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RP STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=1126551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Posfai G., Heckett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouzis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blatner F.R.;
 RT "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7." ;
 RL Nature 409:559-563(2001).
 RN [5]
 RP SEQUENCE FROM N.A.
 RP STRAIN=O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,
 RA Kihara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohaemorrhagic *Escherichia coli*
 O157:H7 and genomic comparison with a laboratory strain K-12." ;
 RL DNA Res. 8:11-22(2001).
 RN [6]
 RP SEQUENCE OF 1-21 FROM N.A.
 RP STRAIN=K12;
 RX MEDLINE=95157246; PubMed=7854117;
 RA Argyvas A., Olsen A., Normark S.;
 RT "Sigma S-dependent growth-phase induction of the *csdB* promoter in
Escherichia coli can be achieved in vivo by sigma 70 in the absence
 of the nucleoid-associated protein H-NS." ;
 RL Mol. Microbiol. 13:11021-10321(1994).
 RN [7]
 RP -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
 CURLIN MONOMERS.
 CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
 CC -----
 CC THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION
 CC BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION-
 CC THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS
 CC USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY
 CC MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL
 CC ENTITIES REQUIRES A LICENSE AGREEMENT (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X90754; CAA62281.1; -

DR EMBL; AE000205; AAC74125.1; -
 DR EMBL; D90741; BAA35831.1; -
 DR EMBL; AE005315; AAC55787.1; -
 DR EMBL; AP002554; BAB34842.1; -
 DR PIR; C90806; C90806.
 DR PIR; G85665; G85665.
 DR PIR; S70787; S70787.
 DR EcoGene; EG12621; csGB.
 KM Fimbria; Signal; Complete proteome.
 FT SIGNAL 1 21
 FT CHAIN 22 151 MINOR CURLIN SUBUNIT.
 FT SEQUENCE 151 AA; 15882 MW; B18D26B9640148 CRC64;
 QY Query Match 12.5%; Score 96.5; DB 1; Length 151;
 Best Local Similarity 28.1%; Pred. No. 0.097;
 Matches 32; Conservative 15; Mismatches 62; Indels 5; Gaps 2;
 Db 21 AAGYDLANSEYNF---AVNELSKSSFNQAAIIGAGTNNSAQLRGSGSKLLAVVAQEGS 76
 QY 38 SSGPDSLTISYQGSNAAALALQDARKYDQVTRVTHMAHAGGADNSTELTONGF 97
 Db 21 AAGYDLANSEYNF---AVNELSKSSFNQAAIIGAGTNNSAQLRGSGSKLLAVVAQEGS 76
 QY 98 RNNATIDONNAKNSDITVQYQGNNAALVNTQASDSVVRQYGFENNAITANQY 151
 Db 77 SNRAKIDGTGYDYNL-AVIDQASANDASISQAGYNTAMIIQSGSNKANTQY 129
 RESULT 5
 ID ICEN_XANCT STANDARD; PRT; 1567 AA.
 AC P18127;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Ice nucleation protein.
 GN INAX.
 OS Xanthomonas campestris (pv. translucens).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 OC Xanthomonadaceae; Xanthomonas.
 OX NCBI_TaxID=343;
 RN [1]
 RP SEQUENCE FROM N.A.
 RP STRAIN=X565;
 RX MEDLINE=91080859; PubMed=2259339;
 RA Zhao J., Orser C.S.;
 RT "Conserved repetition in the ice nucleation gene *inax* from
Xanthomonas campestris pv. *translucens*." ;
 RL Mol. Gen. Genet. 223:163-166(1990).
 CC -1- FUNCTION: Ice nucleation proteins enable bacteria to nucleate
 CC crystallization in supercooled water.
 CC -1- SUBCELLULAR LOCATION: Outer membrane (By similarity).
 CC -1- DOMAIN: CONTAINS 153 IMPERFECT REPEATS OF THE CONSENSUS
 CC OCTAPEPTIDE A-G-Y-G-S-T-L-T; FURTHER ON A 16-RESIDUE AND A
 CC REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.
 CC -1- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE
 CC NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.
 CC -1- SIMILARITY: Belongs to the bacterial ice nucleation protein
 CC family.
 CC -----
 CC THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION
 CC BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION-
 CC THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS
 CC USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY
 CC MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL
 CC ENTITIES REQUIRES A LICENSE AGREEMENT (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X52970; CAA37140.1; -
 DR HSSP; P06620; 11NA.
 DR InterPro; IPR000258; Ice_nucleatn.
 DR Pfam; PF00818; Ice_nucleation; 81.
 DR PRINTS; PR00327; ICENUCLEATN.
 DR PROSITE; PS00314; ICE_NUCLEATION; 57.
 KM Ice nucleation; Repeat; Outer membrane.

between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.emb-sib.ch/announce/> or send an email to license@sib-sib.ch).

CC EMBL; Y00647; CAA68673.1; -

DR PIR; P10221; P10221.

DR PDB; 1LML; 17-SEP-97.

DR MEROPS; M08.001; -

DR InterPro; IPR006025; Pept_M_Zn_BS.

DR InterPro; IPR001577; Peptidase_M8.

DR Pfam; PF01457; Peptidase_M8; 1.

DR PRINTS; PR00782; LSHMANOLYSIN.

DR PROSITE; PS00142; ZINC_PROTEASE; 1.

KW Hydroxylase; Metalloprotease; Glycoprotein; Metal-binding; Zinc; Zymogen; Signal; Cell adhesion; GPI-anchor; 3D-structure; Lipoprotein.

FT SIGNAL 1 39

FT PROPEP 40 100

FT CHAIN 101 577

FT PROPEP 578 602

FT METAL 264 264

FT ACT_SITE 265 265

FT METAL 268 268

FT METAL 334 334

FT DISULFID 125 142

FT DISULFID 191 230

FT DISULFID 314 386

FT DISULFID 393 455

FT DISULFID 406 425

FT DISULFID 415 489

FT DISULFID 466 510

FT DISULFID 515 565

FT DISULFID 535 558

FT CARBOHYD 300 300

FT CARBOHYD 407 407

FT LIPID 577 577

FT STRAND 101 102

FT STRAND 107 108

FT STRAND 111 114

FT HELIX 116 119

FT TURN 121 122

FT TURN 128 129

FT STRAND 131 133

FT STRAND 139 141

FT HELIX 144 146

FT HELIX 150 158

FT TURN 159 159

FT HELIX 160 169

FT TURN 170 171

FT STRAND 172 174

FT STRAND 177 178

FT STRAND 180 181

FT TURN 189 190

FT HELIX 191 193

FT HELIX 198 202

FT TURN 203 203

FT STRAND 205 206

FT STRAND 210 215

FT TURN 221 222

FT STRAND 226 232

FT TURN 234 235

FT STRAND 238 244

FT HELIX 247 249

FT HELIX 256 269

FT TURN 270 271

FT HELIX 274 279

FT TURN 280 281

FT STRAND 283 286

FT HELIX 291 291

FT STRAND 296 299

FT HELIX 302 312

N-LINKED (GLCNAC. . .) (POTENTIAL).

N-LINKED (GLCNAC. . .) (POTENTIAL).

GPI-anchor amidated asparagine.

FT TURN 313 313

FT TURN 315 316

FT STRAND 320 322

FT TURN 328 332

FT STRAND 335 335

FT TURN 337 339

FT TURN 341 342

FT STRAND 343 343

FT STRAND 344 345

FT STRAND 353 353

FT HELIX 356 364

FT TURN 365 366

FT STRAND 372 370

FT HELIX 372 374

FT TURN 380 383

FT HELIX 386 390

FT STRAND 394 395

FT TURN 396 397

FT STRAND 398 399

FT TURN 402 404

FT STRAND 413 414

FT TURN 417 418

FT STRAND 421 425

FT STRAND 428 429

FT HELIX 435 437

FT TURN 443 444

FT STRAND 445 446

FT TURN 450 454

FT STRAND 458 465

FT TURN 466 467

FT HELIX 470 472

FT TURN 475 477

FT HELIX 478 480

FT TURN 485 486

FT STRAND 487 494

FT STRAND 496 496

FT STRAND 506 516

FT TURN 517 520

FT STRAND 521 525

FT TURN 527 528

FT STRAND 533 534

FT TURN 540 542

FT STRAND 543 545

FT HELIX 546 546

FT TURN 550 550

FT STRAND 552 553

FT TURN 555 557

FT STRAND 561 565

FT TURN 566 567

FT HELIX 569 572

FT TURN 573 573

SQ SEQUENCE 602 AA; 63953 MW; 982EF3245D87C43E CRC64;

Query Match 11.8%; Score 91; DB 1; Length 602;

Best Local Similarity 37.5%; Pred. No. 1.4;

Matches 27; Conservative 6; Mismatches 15; Indels 24; Gaps 3;

OY 63 ARKIDQVTVRYTHMAHA---GQADNSTIELTONGFRNNATIDOMNANSITV---115

DB 251 ASRYDQVTVRYTHMAHAALGFSGFPEPARIV-----ANVPVNRGNFDPVPVINS 301

OY 116 -----GOYG 119

DB 302 STAVAKAREQYQ 313

RESULT 8

ID CSGB_SALTI STANDARD; PRT: 151 AA.

AC 0827M3;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Minor curlin subunit precursor.
 CSGB OR STYL180 OR T1777.
 OS *Salmonella typhi*.
 OC Bacteria: Proteobacteria: Gammaproteobacteria: Enterobacteriales;
 CC Enterobacteriaceae: *Salmonella*.
 NX NCBI_TaxID=601;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=CT18;
 RX MEDLINE=21534947; PubMed=11677608;
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
 Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebalina M.,
 Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
 Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
 Felwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jorgels K.,
 Krogh A., Larsen T.S., Leather S., Moule S., O'Gara P., Parry C.,
 Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
 Whitehead S., Barrett B.G.;
 RT "Complete genome sequence of a multiple drug resistant *Salmonella*
 LT enterica serovar Typhi CT18.";
 RL Nature 413:848-852(2001).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN=Ty2 / ATCC 700931;
 RX MEDLINE=22531367; PubMed=12644504;
 RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
 Burland V., Kodoyanni V., Schwartz D.C., Blattner F.R.;
 RT "Comparative genomics of *Salmonella enterica* serovar Typhi strains Ty2
 and CT18.";
 RL J. Bacteriol. 185:2330-2337(2003).
 CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
 CC CURLIN MONOMERS.
 CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>
 CC or send an email to license@1sb-sib.ch).
 CC -----
 DR EMBL: A627269; CAD08267.1; -;
 KW EMBL: AE016840; AA069400.1; -;
 DR Fimbrin; Signal; Complete proteome.
 FT SIGNAL 1 21
 FT CHAIN 22 151 MINOR CURLIN SUBUNIT.
 SO SEQUENCE 151 AA; 16254 MW; 161C54326E573495 CRC64;
 QY Query Match 11.4%; Score 88.5; DB 1; Length 151;
 DB Best Local Similarity 31.0%; Pred. No. 0.48;
 DT Matches 26; Conservative 13; Mismatches 42; Indels 3; Gaps 2;
 DB 69 LVTGVYTHMAHAGGADNSTIELTNGFRNNAITIDONAKXSDIR-VGQYGNNAALVN 127
 DB 48 IIGQVGTDSARVRQGSKLSVTSQEGNNRAKVD--AGNYPAYIEQTGNANDASIS 105
 QY 128 QTPASDSVAVRQVGFNNATANY 151
 DB 106 QSAVGNSSAIIQKSGNKANITQY 129
 DB 106 QSAVGNSSAIIQKSGNKANITQY 129
 RESULT 9
 CSGB SALTY STANDARD; PRT; 151 AA.
 AC PSS236;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Minor curlin subunit precursor (Fimbrin SEF17 minor subunit).
 GN CSGB OR AGFB OR STM1143.
 OS *Salmonella typhimurium*, and
 OS *Salmonella enteritidis*.
 OC Bacteria: Proteobacteria: Gammaproteobacteria: Enterobacteriales;
 CC Enterobacteriaceae: *Salmonella*.
 NX NCBI_TaxID=602, 592;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC SPECIES=S.typhimurium; STRAIN=SR-11;
 RX MEDLINE=98117058; PubMed=9457880;
 RA Rowling U., Bian Z., Hammar M., Sieralta W.D., Normark S.;
 RT "Curli fibers are highly conserved between *Salmonella typhimurium* and
 RL *Escherichia coli* with respect to operon structure and regulation.";
 J. Bacteriol. 180:722-731(1998).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
 RX MEDLINE=21534948; PubMed=11677609;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
 Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
 Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
 Waterston R., Wilson R.K.;
 RT "Complete genome sequence of *Salmonella enterica* serovar Typhimurium
 LT2.";
 RL Nature 413:852-856(2001).
 RN (3)
 RP SEQUENCE FROM N.A.
 RC SPECIES=S.enteritidis; STRAIN=27655-3B;
 RX MEDLINE=96146512; PubMed=8550497;
 RA Collinson S.K., Clouthier S.C., Doran J.L., Baner P.A., Kay W.W.;
 RT "Salmonella enteritidis *agfBAC* operon encoding thin, aggregative
 RL fimbriae.";
 RL J. Bacteriol. 178:662-667(1996).
 CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
 CC CURLIN MONOMERS.
 CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>
 CC or send an email to license@1sb-sib.ch).
 CC -----
 DR EMBL: AJ002301; CA05316.1; -;
 DR EMBL: AE008749; AAL20073.1; -;
 DR PIR: JG6040; JG6040.
 DR StyGene; SG10609; CSGB.
 DR Fimbrin; Signal; Complete proteome.
 FT SIGNAL 1 21
 FT CHAIN 22 151 MINOR CURLIN SUBUNIT.
 SO SEQUENCE 151 AA; 16182 MW; C0FC5430BEDD361D CRC64;
 QY Query Match 11.4%; Score 88.5; DB 1; Length 151;
 DB Best Local Similarity 31.0%; Pred. No. 0.48;
 DT Matches 26; Conservative 13; Mismatches 42; Indels 3; Gaps 2;
 DB 69 LVTGVYTHMAHAGGADNSTIELTNGFRNNAITIDONAKXSDIR-VGQYGNNAALVN 127
 DB 48 IIGQVGTDSARVRQGSKLSVTSQEGNNRAKVD--AGNYPAYIEQTGNANDASIS 105
 QY 128 QTPASDSVAVRQVGFNNATANY 151
 DB 106 QSAVGNSSAIIQKSGNKANITQY 129
 DB 106 QSAVGNSSAIIQKSGNKANITQY 129

RESULT 10
GP63_LEIDO STANDARD; PRT; 590 AA.
ID GP63_LEIDO
AC P23223;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Leishmanolysin precursor (EC 3.4.24.36) (Cell surface protease)
DE Major surface glycoprotein (GP63 protein) (Promastigote surface endopeptidase).
GN GP63.
OS Leishmania donovani.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_Taxid=5661;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LV9;
RX MEDLINE=92107220; PubMed=1762629;
RA Webb J.R., Button L.L., McMaster R.W.;
RT "Heterogeneity of the genes encoding the major surface glycoprotein of Leishmania donovani.";
RL Mol. Biochem. Parasitol. 48:173-184(1991).
CC -1- FUNCTION: Has an integral role during the infection of macrophages in the mammalian host.
CC -1- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and P1' and basic residues at P2 and P3'. A model nonapeptide is cleaved at -Ala-Tyr-[Leu-Lys-Lys-
CC -1- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -1- SIMILARITY: Belongs to peptidase family M8.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M60048; AAA29244.1; -
DR HSSP; P08148; 1LML.
DR MEROPS; M08.001; -
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR InterPro; IPR001577; Peptidase_M8.
DR Pfam; PF01457; Peptidase_M8; 1.
DR PRINTS; PRO0782; LSHMANOLYSIN.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydrolyase; Metalloprotease; Glycoprotein; Metal-binding; Zinc; Zymogen; Signal; Cell adhesion; GPI-anchor; Lipoprotein.
FT SIGNAL 1 39
FT PROPEP 40 87
FT CHAIN 88 565
FT PROPEP 566 590
FT METAL 251 251
FT ACT_SITE 252 252
FT METAL 255 255
FT METAL 321 321
FT DISULFID 112 129
FT DISULFID 178 217
FT DISULFID 301 373
FT DISULFID 380 443
FT DISULFID 393 412
FT DISULFID 402 477
FT DISULFID 454 498
FT DISULFID 503 553
FT DISULFID 523 546
FT CARBOHYD 287 287
FT LIPID 565 565
SQ SEQUENCE 590 AA; 62950 MW; 0FB315D29659F58 CRC64;
Query Match 11.4%; Score 88; DB 1; Length 590;
Best Local Similarity 89.5%; Pred. No. 2.5;

Matches 17; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 63 ARRYDQLVTRVTHENAMA 81
| : |||||
Db 238 ASRYDQLVTRVTHENAMA 256
RESULT 11
GP63_LEICH STANDARD; PRT; 599 AA.
ID GP63_LEICH
AC P15706;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Leishmanolysin precursor (EC 3.4.24.36) (Cell surface protease)
DE Major surface glycoprotein (GP63 protein) (Promastigote surface endopeptidase).
GN GP63.
OS Leishmania chagasi.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_Taxid=44271;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90205976; PubMed=2320059;
RA Miller R.A., Reed S.G., Parsons M.;
RT "Leishmania gp63 molecule implicated in cellular adhesion lacks an Arg-gly-Asp sequence.";
RL Mol. Biochem. Parasitol. 39:267-274(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92112918; PubMed=1370484;
RA Ramamoorthy R., Donelson J.E., Paetz K.E., Maybodi M., Roberts S.C., Wilson M.E.;
RT "Three distinct RNAs for the surface protease gp63 are differentially expressed during development of Leishmania donovani chagasi promastigotes to an infectious form.";
RL J. Biol. Chem. 267:1888-1895(1992).
CC -1- FUNCTION: Has an integral role during the infection of macrophages in the mammalian host.
CC -1- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and P1' and basic residues at P2 and P3'. A model nonapeptide is cleaved at -Ala-Tyr-[Leu-Lys-Lys-
CC -1- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -1- SIMILARITY: Belongs to peptidase family M8.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M60672; AAA29238.1; -
DR EMBL; M28527; AAA29235.1; -
DR PIR; A44951; A44951.
DR HSSP; P08148; 1LML.
DR MEROPS; M08.001; -
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR InterPro; IPR001577; Peptidase_M8.
DR Pfam; PF01457; Peptidase_M8; 1.
DR PRINTS; PRO0782; LSHMANOLYSIN.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydrolyase; Metalloprotease; Glycoprotein; Metal-binding; Zinc; Zymogen; Signal; Cell adhesion; GPI-anchor; Lipoprotein.
FT SIGNAL 1 39
FT PROPEP 40 97
FT CHAIN 98 574
FT PROPEP 575 599
FT METAL 261 261
FT ACT_SITE 262 262
FT METAL 265 265
SQ SEQUENCE 599 AA; 62950 MW; 0FB315D29659F58 CRC64;
Query Match 11.4%; Score 88; DB 1; Length 590;
Best Local Similarity 89.5%; Pred. No. 2.5;

```

FT METAL 331 331 ZINC (CATALYTIC) (BY SIMILARITY).
FT DISULFID 122 139 BY SIMILARITY.
FT DISULFID 188 227 BY SIMILARITY.
FT DISULFID 311 383 BY SIMILARITY.
FT DISULFID 390 452 BY SIMILARITY.
FT DISULFID 403 432 BY SIMILARITY.
FT DISULFID 412 466 BY SIMILARITY.
FT DISULFID 463 507 BY SIMILARITY.
FT DISULFID 512 562 BY SIMILARITY.
FT DISULFID 532 555 BY SIMILARITY.
FT CARBOHYD 297 297 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 394 394 GPI-anchor amidated asparagine (By
FT LIPID 574 574 similarity).
SQ SEQUENCE 599 AA; 63840 MW; 746730AE8E2A2E7C CRC64;

Query Match 11.4%; Score 88; DB 1; Length 599;
Best Local Similarity 89.5%; Pred. No. 2.6;
Matches 17; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

63 ARKQDQLVTRVVTHEMAHA 81
| : |||||
Db 248 ASRYDQLVTRVVTHEMAHA 266

RESULT 12
G63 LEIME STANDARD; PRT; 646 AA.
AC GP63 LEIME
ID GP63 LEIME
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Leishmanolysin C1 precursor (EC 3.4.24.36) (Cell surface protease)
DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface
DE endopeptidase).
GN GP63-C1.
OS Leishmania mexicana.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OC NCBI_TaxID=5665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MNYC/BZ/62/M379;
RX MEDLINE=93149206; PubMed=8426614;
RA Medina-Acosta E., Kares R.E., Russell D.G.;
RT "Structurally distinct genes for the surface protease of Leishmania
RT mexicana are developmentally regulated."
RL Mol. Biochem. Parasitol. 57:31-46(1993).
CC -1- FUNCTION: Has an integral role during the infection of macrophages
CC in the mammalian host.
CC -1- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and
CC P1' and basic residues at P2 and P3'. A model nonapeptide is
CC cleaved at -Ala-Tyr-[Leu-Lys-Lys-".
CC -1- COPACOR: Binds 1 zinc ion per subunit (By similarity).
CC -1- DEVELOPMENTAL STAGE: Expressed in both the promastigote and the
CC amastigote forms.
CC -1- SIMILARITY: Belongs to peptidase family M8.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, X64394, CAA45733.1; -.
CC DR PIR; S19916, S19916.
CC DR HSRP; P08148, 1LML.
CC DR MEROPS; M08.001; -.
CC DR GLYCOSULEDB; P43150; -.
CC DR InterPro; IPR001577; Peptidase_M8.
CC DR Pfam; PF01457; Peptidase_M8; 1.

```

```

DR PRINTS; P00762; LSHMANOLYSIN.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydrolase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;
KW Zymogen; Signal; Cell adhesion; Multigene family.
FT SIGNAL 1 39
FT PROPEP 40 102
FT CHAIN 103 646
FT METAL 266 266 LEISHMANOLYSIN C1.
FT ACT SITE 267 267 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 270 270 BY SIMILARITY.
FT METAL 336 336 ZINC (CATALYTIC) (BY SIMILARITY).
FT DISULFID 127 144 BY SIMILARITY.
FT DISULFID 153 232 BY SIMILARITY.
FT DISULFID 316 388 BY SIMILARITY.
FT DISULFID 395 458 BY SIMILARITY.
FT DISULFID 408 427 BY SIMILARITY.
FT DISULFID 417 492 BY SIMILARITY.
FT DISULFID 469 513 BY SIMILARITY.
FT DISULFID 518 568 BY SIMILARITY.
FT DISULFID 538 561 BY SIMILARITY.
FT CARBOHYD 86 86 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 297 297 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 399 399 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 409 409 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 433 433 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 445 445 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 466 466 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 501 501 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 646 AA; 69054 MW; FE448D8C78C10B0A CRC64;

Query Match 11.4%; Score 88; DB 1; Length 646;
Best Local Similarity 89.5%; Pred. No. 2.8;
Matches 17; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

63 ARKQDQLVTRVVTHEMAHA 81
| : |||||
Db 253 ASRYDQLVTRVVTHEMAHA 271

RESULT 13
PST3 MYCAV STANDARD; PRT; 369 AA.
ID PST3 MYCAV
AC Q9KR89;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Phosphate-binding protein 3 precursor (PBP-3) (Pst3-3).
GN PST3.
OS Mycobacterium avium.
OC Bacteria; Actinobacteria; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OC NCBI_TaxID=1764;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=969A45;
RA Carroll J.D., Wallace R.C., Keane J., Arbelt R.D.;
RT "Identification of Mycobacterium avium DNA sequences that encode
RT exported proteins by using phoB gene fusions."
RL Submitted (MAR-1999) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: Required for binding-protein-mediated phosphate
CC transport (By similarity).
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
CC (By similarity).
CC -1- SIMILARITY: Belongs to the pst3 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

DR EMBL: AF137360; AAF74819.1; -
 DR HSP, P06128; 1A54.
 DR InterPro: IPR000437; Prok_lipprot_s.
 DR InterPro: IPR006059; SLP_bac_1.
 DR Pfam: PF01547; SLP_bac_1; 1.
 DR PROSITE: PS00013; PROKAR_LIPOPROTEIN_1.
 KW Phosphate transport; Transport; Membrane; Lipoprotein; Signal;
 KM Palmitate.
 FT SIGNAL 1 22 POTENTIAL.
 FT CHAIN 23 369 PHOSPHATE-BINDING PROTEIN 3.
 FT LIPID 23 23 N-palmitoyl cysteine (Potential).
 FT LIPID 23 23 S-diacylglycerol cysteine (Potential).
 SQ SEQUENCE 369 AA; 37225 MW; C80EA0AC10F463BC CRC64;

Query Match 11.2%; Score 86.5; DB 1; Length 369;
 Best Local Similarity 27.8%; Pred. No. 2;
 Matches 35; Conservative 21; Mismatches 49; Indels 21; Gaps 6;

QY 1 MLLKVAFAAIVAGSALAGVVPQWGGGNGNGSSGPDSTLSTIYGSANPALALQ 60
 1 MLLNRPFAVLSVLSAGALV-----SGCGSDNNAGAGAAAGSSSKVSCGKKALKASG 54
 DB 61 SDARKYQVLTVV-THMAHAGGADNSTIELTONGFRNNATIDQNNANNSDITVGOY 119
 55 STAQ--NAMTRFVNAFEQACPGQ-----TLNTYANG--SGAGISENGKQTD-----FG 100

QY 120 GNNAL 125
 DB 101 GSDSPL 106

RESULT 14
 FMF2_ECOLI STANDARD; PRT; 188 AA.
 AC P02972;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE F7-2 fimbrial protein precursor (F7-2 pilin).
 GN F7-2 OR PAPA OR C3592.
 OS Escherichia coli O6.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562, 217992;
 RP SEQUENCE FROM N.A.
 RX MEDLINE=8155489; PubMed=6152241;
 RA van Die I., Bergmans H.;
 RT "Nucleotide sequence of the gene encoding the F72 fimbrial subunit of
 a uropathogenic Escherichia coli strain.";
 RL Gene 32:83-90(1984).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92040048; PubMed=1682251;
 RA Denich K., Blyn L.B., Craiu A., Braaten B.A., Hardy J., Low D.A.,
 RA O'Hanley P.D.;
 RT "DNA sequences of three ppa genes from uropathogenic Escherichia
 coli strains: evidence of structural and serological conservation.";
 RL Infect. Immun. 59:3849-3858(1991).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX STRAIN=06:H1 / CFT073 / ATCC 700928;
 RA MEDLINE=22388234; PubMed=12471157;
 RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
 RA Rasko D., Buckles E.L., Lion S.R., Boutin A., Hackett J., Stroud D.,
 RA Mayhew G.F., Rose D.U., Zhou S., Schwartz F.C., Perna N.T.,
 RA Mobley H.L.T., Domenech M.S., Blattner F.R.;
 RT "Extensive mosaic structure revealed by the complete genome sequence
 of uropathogenic Escherichia coli.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
 CC -1- FUNCTION: Fimbriae (also called pili), polar filaments radiating
 from the surface of the bacterium to a length of 0.5-1.5

CC micrometers and numbering 100-300 per cell, enable bacteria to
 CC colonize the epithelium of specific host organs.
 CC -1- SIMILARITY: BELONGS TO THE FIMA/PAPA FAMILY OF FIMBRIA PROTEINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M12861; AAA23778.1; -
 DR EMBL: M68060; AAA24278.1; -
 DR EMBL: AE016766; AAN82040.1; ALT_INIT.
 DR PIR: A03496; YOECP2.
 DR InterPro: IPR008966; Adhes_bact.
 DR InterPro: IPR00259; Fimbrial.
 DR Pfam: PR00419; Fimbrial; 1.
 KW Fimbria; Signal; Complete proteome.
 FT SIGNAL 1 21
 FT CHAIN 22 188 F7-2 FIMBRIAL PROTEIN.
 FT DISULFID 43 82 PROBABLE.
 SQ SEQUENCE 188 AA; 19184 MW; 0BEF750CFD843157 CRC64;

Query Match 10.9%; Score 84.5; DB 1; Length 188;
 Best Local Similarity 25.0%; Pred. No. 1.4;
 Matches 45; Conservative 26; Mismatches 76; Indels 33; Gaps 8;

QY 1 MLLKVAFAAIVAGSALAGVVPQWGGGNGNGSSGPDSTLSTIYGS 52
 2 IKSVAAGAVAAVVSFGAAPTIPQSGKTFNGTVDAFCGIDAGSADSIDPGV-- 59
 DB 53 ANAALQSDA---RKVD-QLVTRVTH-EMAHAGGADNSTIELTONGFRNNATIDQW 106
 60 --SKLFLENDSQPKSPFDILINDITNFKKAAGGAGKATGVTLPFGVPSGQSDML 117
 QY 107 --NAKNSDITVGOYGGNNAALVNOGASDSSVM-----YRQVGGNNAATNQY 151
 118 QTVGATNTAIVTDPHGKRVKFDGATATGVSILVDGNTHTFTAAVRKDSGNSPVTSAGF 177

RESULT 15
 ICEN PANAN STANDARD; PRT; 1034 AA.
 AC Q47879;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Ice nucleation protein inau.
 GN INAU.
 OS Pantoea ananae (Erwinia uredovora).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Pantoea.
 OX NCBI_TaxID=553;
 RP SEQUENCE FROM N.A.
 RX STRAIN=KUN-3;
 RX MEDLINE=94264407; PubMed=7764866;
 RA Michigami Y., Watabe S., Abe K., Obata H., Arai S.;
 RT "Cloning and sequencing of an ice nucleation active gene of Erwinia
 uredovora.";
 RL Biosci. Biotechnol. Biochem. 58:762-764(1994).
 CC -1- FUNCTION: Ice nucleation proteins enable bacteria to nucleate
 CC crystallization in supercooled water.
 CC -1- SUBCELLULAR LOCATION: Outer membrane.
 CC -1- DOMAIN: CONTAINS IMPERFECT REPEATS OF A CONSENSUS OCTAPEPTIDE
 CC A-G-Y-G-S-T-X-T, FURTHER ON A 16-RESIDUE AND A REGIONAL 48-RESIDUE
 CC PERIODICITY IS SUPERIMPOSED.
 CC MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE
 CC NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.
 CC -1- SIMILARITY: Belongs to the bacterial ice nucleation protein
 CC family.

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:23:54 ; Search time 30.5 Seconds
(without alignments)
1562.074 Million cell updates/sec

Title: US-09-543-407-20

Perfect score: 774
Sequence: 1 MLLKVAAPAAIIVSGSALA.....DSVMVROYFGNNATANY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL_25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp Vertebrate:*
- 14: sp Unclassified:*
- 15: sp_rv1rus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	680	87.9	152	2	033802 salmonella
2	582.5	75.3	150	2	07X243 citrobacter
3	552	71.3	149	2	07X240 citrobacter
4	503.5	65.1	152	16	08CW63 escherichia
5	430.5	55.6	150	2	07X237 enterobacte
6	301	38.9	76	2	054069 salmonella
7	122	15.8	29	2	098375 escherichia
8	115	14.9	130	16	08B314 bradyrhizob
9	110	14.2	139	16	08E1H3 shewanella
10	106.5	13.8	502	16	08B1H4 shewanella
11	105.5	13.6	145	16	08U6N9 agrobacteri
12	105.5	13.6	153	16	08B3J6 bradyrhizob
13	99	12.8	573	10	098AF2 arabidopsis
14	98.5	12.7	154	16	08B3T5 bradyrhizob
15	97.5	12.6	1765	16	07V855 prochloroc
16	97	12.5	157	16	08BHG0 pseudomonas

17	96.5	12.5	151	16	07UC21 shigella fl
18	96.5	12.5	160	16	08CW64 escherichia
19	96.5	12.5	160	16	08B3U7 shigella fl
20	96	12.4	2174	16	092U8 rhizobium m
21	95	12.3	1422	16	08BPU3 shewanella
22	94.5	12.2	438	16	082M56 streptomyce
23	94.5	12.2	598	5	025275 leishmania
24	94.5	12.2	1209	16	089CK5 bradyrhizob
25	93	12.0	480	16	089EV2 bradyrhizob
26	92.5	12.0	91	2	09S3J8 escherichia
27	92.5	12.0	151	2	07X244 citrobacter
28	92	11.9	644	5	043994 leishmania
29	92	11.9	1408	16	08B833 shewanella
30	91	11.8	645	16	07U1C5 mycobacteri
31	91	11.8	646	16	053818 mycobacteri
32	91	11.8	1410	16	08CMJ0 shewanella
33	91	11.8	2734	16	089C73 bradyrhizob
34	90.5	11.7	329	3	08N1Z1 neopspora
35	90	11.6	152	2	07X241 citrobacter
36	90	11.6	639	5	025274 leishmania
37	90	11.6	644	5	08MNY9 leishmania
38	90	11.6	714	16	07U5X6 synechococc
39	90	11.6	1460	16	08K6R1 streptococc
40	90	11.6	3056	16	07USQ0 rhodospirell
41	90	11.6	3501	16	08Y106 ralsstonia s
42	90	11.6	3552	16	08XSD6 enterobacte
43	89.5	11.6	151	2	07X238 enterobacte
44	89.5	11.6	297	2	09Z1Y5 borrelia bu
45	89.5	11.6	333	5	08T986 drosophila

ALIGNMENTS

RESULT 1

033802 ID 033802 PRELIMINARY; PRT; 152 AA.

AC 033802; 01-JAN-1998 (TREMBLrel. 05, Created)

DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE Agfa protein (Fragment).

GN AGFA.

OS Salmonella typhimurium.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Salmonella.

OX NCBI_Taxid=602;

RN (1)

RP SEQUENCE FROM N.A.

RX MEDLINE=98053981; PubMed=9393832;

RA Sukupolvi S.S., Lorentz R.G., Gordon J.I., Bian Z., Pfeiffer J.D.,

RA Normark S.J., Rhen M.;

RT "Expression of thin, aggregative fimbriae promotes interaction of

RT Salmonella typhimurium SR-11 with mouse small intestinal epithelial

RT cells."

RL Infect. Immun. 65:5320-5325(1997).

DR EMBL; AJ000514; CAA04151.1; -.

FT NON TER 152 152

SQ SEQUENCE 152 AA; 15401 MW; 9DA7DADC2364B006 CRC64;

Query Match 87.9%; Score 680; DB 2; Length 152;
Best Local Similarity 89.4%; Pred. No. 2.3e-48;
Matches 135; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

QY	1	MLLKVAAPAAIIVSGSALA	60
QY	1	MLLKVAAPAAIIVSGSALA	60
QY	61	SPARKYDOLVTRVVTTHMAHAGAGADNSTIEITONGFRNNATIDOWNAKNSDITVGOYGG	120
QY	61	SPARKSETTITTSVGNGAGDVAGADNSTIEITONGFRNNATIDOWNAKNSDITVGOYGG	120
QY	121	NNALVNTASDSVMVROYFGNNATANY	151

```

Db      121 NNAALVNOTASDSSVWVROVGFNNATPANOY 151
      |||
RESULT 2
ID      Q7X243      PRELIMINARY;      PRT;      150 AA.
AC      Q7X243;
DT      01-OCT-2003 (TREMBLrel. 25, Created)
DT      01-OCT-2003 (TREMBLrel. 25, Last sequence update)
DT      01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE      Curlin-csgA protein.
GN      CSGA.
OS      Citrobacter sp. fec2.
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC      Enterobacteriaceae; Citrobacter.
OX      NCBI_TaxID=213763;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=fec2;
RA      Zogaj X., Bokranz W., Nimtz M., Romling U.;
RT      "Production of Cellulose and Curli Fimbriae by Members of the Family
RT      Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL      Infect. Immun. 72:4151-4158(2003).
DR      EMBL; AJ515700; CAD56672.1; -.
SQ      SEQUENCE 150 AA; 15016 MW; 1D7141B8D6973DC6 CRC64;

Query Match      75.3%; Score 582.5; DB 2; Length 150;
Best Local Similarity 78.8%; Pred. No. 2.3e-40;
Matches 119; Conservative 10; Mismatches 21; Indels 1; Gaps 1;

QY      1 MLLKVAFAAIVVSGSALAGVVPOMGGGNHNGSGSPDSTLSIYQGSANALALQ 60
Db      1 MLLKVAFAAIVVSGSALAGVVPOMGGGNHNGSGSPDSTLSIYQGSANALALQ 59
QY      61 SDARKYDQVTRVVTHEMAHAGGADNSTIELTONGFRNNATTIDQWAKNSDITVGGYGG 120
Db      60 SDARKSDTTHQNGFGNGADVGQGSDDSTIDLTORGFGNSATIDQWNGKSDITVSGYGG 119

QY      121 NNAALVNOTASDSSVWVROVGFNNATPANOY 151
Db      120 HNAALVNOTASDSSVWVROVGFNNATPANOY 150

RESULT 3
ID      Q7X240      PRELIMINARY;      PRT;      149 AA.
AC      Q7X240;
DT      01-OCT-2003 (TREMBLrel. 25, Created)
DT      01-OCT-2003 (TREMBLrel. 25, Last sequence update)
DT      01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE      Curlin-csgA protein.
GN      CSGA.
OS      Citrobacter freundii.
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC      Enterobacteriaceae; Citrobacter.
OX      NCBI_TaxID=546;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=fec4;
RA      Zogaj X., Bokranz W., Nimtz M., Romling U.;
RT      "Production of Cellulose and Curli Fimbriae by Members of the Family
RT      Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL      Infect. Immun. 72:4151-4158(2003).
DR      EMBL; AJ515701; CAD56675.1; -.
SQ      SEQUENCE 149 AA; 15260 MW; 946DD52017F648FD CRC64;

Query Match      71.3%; Score 552; DB 2; Length 149;
Best Local Similarity 73.5%; Pred. No. 7.2e-38;
Matches 111; Conservative 18; Mismatches 20; Indels 2; Gaps 1;

QY      1 MLLKVAFAAIVVSGSALAGVVPOMGGGNHNGSGSPDSTLSIYQGSANALALQ 60

```

```

Db      1 MLLKVAFAAIVVSGSALAGVVPOM--GGNHGSGSNVCPDSSLSIYQGSNNSANALQ 58
QY      61 SDARKYDQVTRVVTHEMAHAGGADNSTIELTONGFRNNATTIDQWAKNSDITVGGYGG 120
Db      59 SDARKSDVITTHQNGFGNGADVGQGSDDSTLSIQTGFGNSATIDQWAKNADISVTVQFG 118

QY      121 NNAALVNOTASDSSVWVROVGFNNATPANOY 151
Db      119 RAGALVNOTASDSSVWVROVGFNNATPANOY 149

RESULT 4
ID      O8CW63      PRELIMINARY;      PRT;      152 AA.
AC      O8CW63;
DT      01-MAR-2003 (TREMBLrel. 23, Created)
DT      01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT      01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE      Major curlin subunit precursor.
GN      CSGA OR C1306.
OS      Escherichia coli O6.
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC      Enterobacteriaceae; Escherichia.
OX      NCBI_TaxID=217992;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=O6:H / CFT073 / ATCC 700928;
RX      MEDLINE=22388234; PubMed=12471157;
RA      Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA      Raske D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA      Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA      Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT      "Extensive mosaic structure revealed by the complete genome sequence
RT      of uropathogenic Escherichia coli.";
RL      Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
DR      EMBL; AB016759; AAN79779.1; -.
KW      Complete proteome.
SQ      SEQUENCE 152 AA; 15064 MW; 3BA57F34C1240EB3 CRC64;

Query Match      65.1%; Score 503.5; DB 16; Length 152;
Best Local Similarity 67.1%; Pred. No. 7.1e-34;
Matches 102; Conservative 20; Mismatches 29; Indels 1; Gaps 1;

QY      1 MLLKVAFAAIVVSGSALAGVVPOM--GGGNHNGSGNSGPDSTLSIYQGSANALALQ 59
Db      1 MLLKVAFAAIVVSGSALAGVVPYGGGGNHGSGNSGNSLNTIYQGGNSALALQ 60
QY      60 QSDARKYDQVTRVVTHEMAHAGGADNSTIELTONGFRNNATTIDQWAKNSDITVGGYGG 119
Db      61 QADARNSDLTTHQHGNGGADVGQGSDDSSIDLTOGRFGNSATIDQWNGKSDITVQKQFG 120

QY      120 GNNALVNOTASDSSVWVROVGFNNATPANOY 151
Db      121 GAGGAADVDTIASNSSVWVROVGFNNATPANOY 152

RESULT 5
ID      Q7X237      PRELIMINARY;      PRT;      150 AA.
AC      Q7X237;
DT      01-OCT-2003 (TREMBLrel. 25, Created)
DT      01-OCT-2003 (TREMBLrel. 25, Last sequence update)
DT      01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE      Curlin-csgA protein.
GN      CSGA.
OS      Enterobacter sakazakii.
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC      Enterobacteriaceae; Enterobacter.
OX      NCBI_TaxID=28141;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=fec39;
RA      Zogaj X., Bokranz W., Nimtz M., Romling U.;

```


DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
 DE Minor curlin subunit CsgB, putative.
 GN S00866
 OS Shewanella oneidensis.
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
 OC Alteromonadaceae; Shewanella.
 CX NCBI_TaxID=70863;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MR-1;
 RX MEDLINE=22297686; PubMed=12368813;
 RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
 Read T.D., Eissen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
 Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
 Deboy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
 Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
 Vamathevan J., Weidman J., Imprim M., Lee K., Berry K., Lee C.,
 Mueller J., Khouri H., Gill J., Ullrichback T.R., McDonald L.A.,
 RA Feldblyum T.V., Smith H.O., Venter J.C., Nealon K.H., Fraser C.M.,
 RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
 RT Shewanella oneidensis."
 RL Nat. Biotechnol. 20:1118-1123(2002).
 DR EMBL; AEO15532; AAN53941.1; -
 DR TIGR; S00866; -
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 502 AA; 52441 MW; D08CA23D6C46B6ZD CRC64;

Query Match 14.2%; Score 110; DB 16; Length 139;
 Best Local Similarity 28.3%; Pred. No. 0.13; Mismatches 34; Indels 24; Gaps 3;
 Matches 30; Conservative 18; Mismatches 34; Indels 24; Gaps 3;

QY 39 SGPDSSTIYOGSANAALALQSDARKYDQVTRVTHHEMAHQAGADNSTIELTONGFR 98
 DB 41 SGRNLIDIVGQGGANGIVFGS-----GSDNAX-VTQAGND 77
 QY 99 NNATIDQNAKSDITVGOYGNNALVNOTASDSSVMRYGFGNATANOY 151
 DB 78 NISLVTOIGT-NNEVOLLQVGAQNKASITQIGNDLVOLNQLSGSN 122

RESULT 10
 Q8E1H4 PRELIMINARY; PRT; 502 AA.
 AC Q8E1H4;
 DT 01-MAR-2003 (TReMBLrel. 23, Created)
 DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
 DE Conserved hypothetical protein.
 GN S00865.
 OS Shewanella oneidensis.
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
 OC Alteromonadaceae; Shewanella.
 CX NCBI_TaxID=70863;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MR-1;
 RX MEDLINE=22297686; PubMed=12368813;
 RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
 Read T.D., Eissen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
 Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
 Deboy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
 Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
 Vamathevan J., Weidman J., Imprim M., Lee K., Berry K., Lee C.,
 Mueller J., Khouri H., Gill J., Ullrichback T.R., McDonald L.A.,
 RA Feldblyum T.V., Smith H.O., Venter J.C., Nealon K.H., Fraser C.M.,
 RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
 RT Shewanella oneidensis."
 RL Nat. Biotechnol. 20:1118-1123(2002).
 DR EMBL; AEO15532; AAN53941.1; -
 DR TIGR; S00866; -
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 502 AA; 52441 MW; D08CA23D6C46B6ZD CRC64;

Query Match 13.8%; Score 106.5; DB 16; Length 502;
 Best Local Similarity 23.1%; Pred. No. 1.2;
 Matches 40; Conservative 30; Mismatches 72; Indels 31; Gaps 5;

QY 10 AAIVSGSALAGVPPWGGG-----GNHNG-----GNS-----SGPD 42
 DB 163 AVPRVEGDNDGDGIKQYAGNNNOAGLIALDLASAVGNNNDVSRQIINNRFKAQKJAGND 222
 QY 43 STLSIYOGSANAALALQSDARKYDQVTRVTHHEMAHQAGADNSTIELTONGFR-- 99
 DB 223 NSVDIYOKDQNHGTFYVALAGSENDISMEQSGSNNTAYLSMTTGDNTVDITQDGSNTV 282
 QY 100 -NATIDQNAKSDITVGOYGNNALVNOTASDSSVMRYGFGNATANOY 151
 DB 283 GBSLIDIGDGDNDITIKQKQDNGAEPFWDGSDNVDLKQGDANFAIFGAY 335

RESULT 11
 Q8U6N9 PRELIMINARY; PRT; 145 AA.
 ID Q8U6N9;
 AC Q8U6N9;
 DT 01-JUN-2002 (TReMBLrel. 21, Created)
 DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
 DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
 DE Hypothetical protein Atu4768.
 GN ATU4768 OR AGR L 228.
 OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
 CC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
 CX NCBI_TaxID=176299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21608550; PubMed=11743193;
 RA Wood D.W., Serrhal J.C., Kaul R., Monks D.E., Kitajima J.P.,
 Ockra V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Jr., Woo L.,
 Chen Y., Paulsen I.T., Eissen J.A., Karp P.D., Bovee D. Sr.,
 Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
 RA Kutayavin T., Levy R., Li M.-J., McClelland E., Palmeri A.,
 Raymond C., Rouse G., Saenphimachak C., Wu Z., Romero P., Gordon D.,
 Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
 RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-X., Dolan M.,
 RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
 RA Nester E.W.;
 RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
 RT C58."
 RL Science 294:2317-2323(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21608551; PubMed=11743194;
 RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
 RA Guroilo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
 RA Houtmel K., Gordon J., Vaudin M., Iarchouk O., Epp A., Liu F.,
 RA Wollam C., Allinger M., Dougherty D., Scott C., Jappas C., Markelz B.,
 RA Flanagan C., Crowell C., Gursan J., Lomo C., Sear C., Strub G.,
 RA Cielo C., Slater S.;
 RT "Genome sequence of the plant pathogen and biotechnology agent
 RT Agrobacterium tumefaciens C58."
 RL Science 294:2323-2328(2001).
 DR EMBL; AEO09405; AAL45562.1; -
 DR EMBL; AEO08209; AAK88682.1; -
 DR PIR; AD3143; AD3143.
 DR PIR; H98144; H98144.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 145 AA; 14984 MW; DEDC870E1713D51A CRC64;

Query Match 13.6%; Score 105.5; DB 16; Length 145;
 Best Local Similarity 23.2%; Pred. No. 0.32;
 Matches 36; Conservative 27; Mismatches 61; Indels 31; Gaps 4;

QY 3 LKVAAPAIIVYSGSALAGVPPWGG-----GNHNGGNSGPDSTLSIYOGSANA 55
 DB 1 MIRKSPIASALVALVGLSAAAPAMANDVRIEYQWNSAGAGQEGYGNRIRITYONGGYNR 60

QY 56 ALALQDARKYDQVTRVYTHEMAHAGGADNSTIELTONGFRNNATIDQMNKNSDIY 115
 DB 61 IVHQ-----YGC-----HNTSAYGQEGHDSYGTONGNHNVAGI----- 96
 QY 116 GQYGNNAALVNQTSADSSVMVRQVGFNNATANO 150
 DB 97 GQYGNHTTILITDQNGNNTAGVQVGRGGSANVSQ 131

RESULT 12

Q89J16 PRELIMINARY; PRT; 153 AA.
 AC Q89J16;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DE 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DB C89B proteoln.
 GN C89B OR BLS5297.
 OS Bradyrhizobium japonicum.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Bradyrhizobiaceae; Bradyrhizobium.
 OX NCBI_Taxid=375;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=USDA 110;
 RX MEDLINE=2484998; PubMed=12597275;
 RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
 RA Sasamoto S., Matanabe A., Idesawa K., Iriuchi M., Kawashima K.,
 RA Kohara M., Matsumoto M., Shimo S., Tsuruoka H., Wada T., Yamada M.,
 RA Tabata S.;
 RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
 Bradyrhizobium japonicum USDA110."
 RL DNA Res. 9:189-197(2002).
 DR EMBL; AP005954; BAC50562.1; -;
 KW Complete proteome.
 SQ SEQUENCE 153 AA; 15991 MW; 4CE71DEAC375145B CRC64;

Query Match 13.6%; Score 105.5; DB 16; Length 153;
 Best Local Similarity 25.7%; Pred. No. 0.34;
 Matches 39; Conservative 35; Mismatches 55; Indels 23; Gaps 7;

QY 2 KLLKVA--FAAIYSSALAGVPPWGGGNNHGGNSSGPDSTL-SIYOYSANALAL 59
 DB 10 RVLAAVALLAAGAAQASGSIQR-----SYNPNPISITIVYFGNDVQPPYI 58
 QY 60 OSDARKYDQVTRVYTHEMAHAGGADNSTIELTONGFRNNATIDQMNKNSDIYGOYX 119
 DB 59 EHSKRN---IARV-----QIGSSGVDAFI--IQNGTRYANAVIQMG-TTNAVQSGS 108
 QY 120 GNNALVNQTSADSSVMVRQVGFNNATANOY 151
 DB 109 LSNVADITQIGNSTNALLQIGMNSGAVRQF 140

RESULT 13

Q89AF2 PRELIMINARY; PRT; 573 AA.
 AC Q89AF2;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 25, Last sequence update)
 DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE F3F19.21 protein (Hypothetical protein).
 DE F3F19.21 OR ATG13190, F3F19.21 OR ATG13190.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC euroside 11; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_Taxid=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RA Vreotsekaia V.S., Schwartz J.R., Yu G., Toriumi M., Lee J.M., Lenz C.,
 RA Liu S., Li J., Kremenetskaia I., Luoro J., Ngan I., Gonzalez A.,

Altai H., Araujo R., Chao Q., Conn L., Conway A.B., Dunn P.,
 Hansen N., Hultzer L., Kim C., Palm C., Rowley D., Shim P., Walker M.,
 Davis R.W., Ecker J., Federisiel N.A., Theologis A.;
 RT "Arabidopsis thaliana chromosome 1 BAC F3F19 sequence."
 RN Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RA Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Southwick A., Katlin-Neumann G., Nguyen M., Lam B., Miranda M.,
 RA Palm C.J., Bowser L., Jones T., Banh J., Carinini P., Chen H.,
 RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A.,
 RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakurai H.,
 RA Sakurai T., Satou M., Seki M., Shim P., Yamada K., Shinozaki K.,
 RA Ecker J., Theologis A., Davis R.W.;
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Nguyen M., Southwick A., Tripp M., Palm C.J., Jones T., Wu T.,
 RA Carinini P., Chen H., Cheuk R., Chan M.M., Chang C.H., Dale J.M.,
 RA Deng J.M., Hayashizaki Y., Hsuan V.W., Lee J.M., Ishida J., Kamiya A.,
 RA Kawai J., Kim C.J., Narusaka M., Onodera C.S., Quach H.L., Sakurai T.,
 RA Satou M., Seki M., Shim P., Tang C.C., Toriumi M., Wong C., Wu H.C.,
 RA Yamada K., Yu G., Yuan S., Shinozaki K., Ecker J., Theologis A.,
 RA Davis R.W.;
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC007357; AAL31072.1; -;
 DR EMBL; AY062527; AAL32605.1; -;
 DR EMBL; BT002575; AAO00935.1; -;
 DR PIR; C86266; C86266.
 DR GO; GO:0003676; P:nucleic acid binding; IEA.
 DR InterPro; IPR000504; RNA_rec_mot.
 DR Pfam; PF00076; RRM; 1.
 DR SMART; SM00360; RRM; 1.
 DR PROSITE; PS50102; RRM; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 573 AA; 61400 MW; 0D1C10384B86C2BC CRC64;

Query Match 12.8%; Score 99; DB 10; Length 573;
 Best Local Similarity 25.4%; Pred. No. 5.9;
 Matches 34; Conservative 24; Mismatches 46; Indels 30; Gaps 6;

QY 13 VVSGSALAGVPPWGGGNNHGGNSSGPDSTLSIYOYSANALALQDARKYDQVTR 72
 DB 84 IYSGGIVEG---KYRNDGHNH---ISGPDTRSDVYPPQASSFGAGLNDI----- 127
 QY 73 VVTHEMAHAGGADNSTIELTONGFRNNA--TIDQMNKNS-----DITVGOYGN 122
 DB 128 IQSNKLAQGG---STTVVLNHNHGFSGNNAVNPENKPVHNSYGAPGAGQOIPVSGQSVNP 183
 QY 123 AALVNQTSADSSVM 136
 DB 184 NYMMNKSPTQSEFV 197

RESULT 14

Q89J15 PRELIMINARY; PRT; 154 AA.
 AC Q89J15;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DE 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE BLS5298 protein.
 GN BLS5298.
 OS Bradyrhizobium japonicum.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Bradyrhizobiaceae; Bradyrhizobium.
 OX NCBI_Taxid=375;
 RN [1]
 RP SEQUENCE FROM N.A.

Altai H., Araujo R., Chao Q., Conn L., Conway A.B., Dunn P.,
 Hansen N., Hultzer L., Kim C., Palm C., Rowley D., Shim P., Walker M.,
 Davis R.W., Ecker J., Federisiel N.A., Theologis A.;
 RT "Arabidopsis thaliana chromosome 1 BAC F3F19 sequence."
 RN Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RA Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Southwick A., Katlin-Neumann G., Nguyen M., Lam B., Miranda M.,
 RA Palm C.J., Bowser L., Jones T., Banh J., Carinini P., Chen H.,
 RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A.,
 RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakurai H.,
 RA Sakurai T., Satou M., Seki M., Shim P., Yamada K., Shinozaki K.,
 RA Ecker J., Theologis A., Davis R.W.;
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Nguyen M., Southwick A., Tripp M., Palm C.J., Jones T., Wu T.,
 RA Carinini P., Chen H., Cheuk R., Chan M.M., Chang C.H., Dale J.M.,
 RA Deng J.M., Hayashizaki Y., Hsuan V.W., Lee J.M., Ishida J., Kamiya A.,
 RA Kawai J., Kim C.J., Narusaka M., Onodera C.S., Quach H.L., Sakurai T.,
 RA Satou M., Seki M., Shim P., Tang C.C., Toriumi M., Wong C., Wu H.C.,
 RA Yamada K., Yu G., Yuan S., Shinozaki K., Ecker J., Theologis A.,
 RA Davis R.W.;
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC007357; AAL31072.1; -;
 DR EMBL; AY062527; AAL32605.1; -;
 DR EMBL; BT002575; AAO00935.1; -;
 DR PIR; C86266; C86266.
 DR GO; GO:0003676; P:nucleic acid binding; IEA.
 DR InterPro; IPR000504; RNA_rec_mot.
 DR Pfam; PF00076; RRM; 1.
 DR SMART; SM00360; RRM; 1.
 DR PROSITE; PS50102; RRM; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 573 AA; 61400 MW; 0D1C10384B86C2BC CRC64;

RC STRAIN=USDA 110;
 RX MEDLINE=22484998; PubMed=12597275;
 RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiyumi T.,
 RA Sasamoto S., Watanabe A., Idegawa K., Iriyuchi M., Kawashima K.,
 RA Kohara M., Matsumoto M., Shimo S., Tsuruoka H., Wada T., Yamada M.,
 RA Tabata S.;
 RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
 RT Bradyrhizobium japonicum USDA110.";
 RL DNA Res. 9:189-197(2002).
 DR EMBL; AP005954; BAC50563.1; -.
 KM Complete proteome.
 SQ SEQUENCE 154 AA; 15420 MW; 9D698D711E2ED19 CRC64;

Query Match 12.7%; Score 98.5; DB 16; Length 154;
 Best Local Similarity 36.8%; Pred. No. 1.3;
 Matches 25; Conservative 12; Mismatches 28; Indels 3; Gaps 2;

QY 84 GADNSTELTQNGFRNNATIDQWNAKNSDITVGGYGGNNAALVNTQASDSSVMVRYQVFG 143
 DB 77 GSTSAIV--VONGINMATGVILQFGGINS-ASVQAGMNNFAFVGQTGSATSLVSQLGAM 133

QY 144 NNATANQY 151
 DB 134 NTGTVAQF 141

RESULT 15

Q7V8S5 PRELIMINARY; PRT; 1765 AA.

AC Q7V8S5;

DT 01-OCT-2003 (TREMBlrel. 25, Created)

DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)

DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)

DE Hemolysin-type calcium-binding region: RTX N-terminal domain.

GN PMT0256.

OS Prochlorococcus marinus (strain MIT 9313).

OC Bacteria; Cyanobacteria; Prochlorophytes; Prochlorococcaceae;

OC Prochlorococcus.

OC NCBI_TaxID=74547;

CK [1]

RN SEQUENCE FROM N.A.

RP MEDLINE=22825698; PubMed=12917642;

RA Rocap G., Larimer F.W., Lamerdin J., Malfatti S., Chain P.,

RA Ahlgren N.A., Arellano A., Coleman M., Hauser L., Hess W.R.,

RA Johnson Z.I., Land M., Lindell D., Post A.F., Regala W., Shah M.,

RA Shaw S.L., Steglich C., Sullivan M.B., Ting C.S., Tolonen A.,

RA Webb E.A., Zinser E.R., Chisholm S.W.;

RT "Genome divergence in two Prochlorococcus ecotypes reflects oceanic

RT niche differentiation.";

RL Nature 424:1042-1047(2003).

DR EMBL; BX572095; CAE20431.1; -.

KM Complete proteome.

SO SEQUENCE 1765 AA; 187603 MW; 817CE4F5007580CD CRC64;

Query Match 12.6%; Score 97.5; DB 16; Length 1765;

Best Local Similarity 28.7%; Pred. No. 31;

Matches 41; Conservative 21; Mismatches 50; Indels 31; Gaps 8;

QY 5 KYAAPAA-----IVSSGALAGVPPQWGGGNNNGGSSGPDSTLSIYQY 51
 DB 222 KYAAAYGALNSDITQSGDDFIYNAS-SNRGWMRGGYAVAGVANG-----SSINTG 275
 QY 52 SANAAALU--OSDARKYDQLVTRVYTHEVAHAGAGANSTIELTQNGFRNNATID-QWNA 108
 DB 276 DGNDAIHITARDGRTTNAMAR--DSNINTSGSDSVTL-----NAFTNSRFIDPAYGA 328
 QY 109 KNSDITVGGYGGNNAALVNTQAS 131
 DB 329 SNSINLG--SGNDSLILINASAN 349

Search completed: March 11, 2004, 18:40:42
 Job time : 39.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:13:53 ; Search time 45.9 Seconds
(without alignments)
929.514 Million cell updates/sec

Title: US-09-543-407-22

Perfect score: 776
Sequence: 1 MKLLKVAAPPAIVSGSALA.....DSSVMVROYFGNNATANY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_297an04:*
1: geneseqp19808:*
2: geneseqp19908:*
3: geneseqp20008:*
4: geneseqp20018:*
5: geneseqp20028:*
6: geneseqp20038:*
7: geneseqp20038:*
8: geneseqp20048:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	776	100.0	151	3	AAB36351 AgfA::PT3
2	712	91.8	151	3	AAB36354 AgfA::PT3
3	682	87.9	151	2	AAR74625 AgfA sequ
4	682	87.9	151	3	AAB36341
5	677	87.2	151	2	AAB23570
6	657	84.7	151	3	AAB36355
7	619	79.8	151	3	AAB36350
8	609	78.5	151	3	AAB36352
9	604	77.8	151	3	AAB36346
10	602	77.6	151	3	AAB36347
11	601	77.4	151	3	AAB36353
12	600	77.3	151	3	AAB36349
13	567	73.1	151	3	AAB36348
14	521	67.1	151	3	AAB36343
15	516	66.5	151	7	ABR82651 E. coli C
16	497	64.0	120	2	AAR62761
17	497	64.0	120	2	AAR23569
18	443	57.1	142	2	AAR52664
19	371	47.8	122	2	AAR52663
20	146	18.8	45	3	AAB36316
21	132	17.0	22	3	AAB36318
22	115	14.8	22	3	AAB36325
23	115	14.8	22	3	AAB36339
24	115	14.8	22	3	AAB36320
25	113	14.6	24	7	ABR82644 E. coli C

26	111	14.3	22	3	AAB36322	Aab36322 Salmonella
27	111	14.3	22	3	AAB36327	Aab36327 Salmonella
28	111	14.3	22	3	AAB36337	Aab36337 Salmonella
29	109	14.0	23	3	AAB36340	Aab36340 Salmonella
30	109	14.0	23	3	AAB36324	Aab36324 Salmonella
31	109	14.0	23	3	AAB36319	Aab36319 Salmonella
32	102	13.1	26	7	ABR82649	ABR82649 E. coli V
33	98	12.6	26	7	ABR82645	ABR82645 E. coli C
34	97.5	12.6	520	6	AAO16497	AAO16497 Arg10pe t
35	96	12.4	19	3	AAB36323	Aab36323 Salmonella
36	96	12.4	19	3	AAB36336	Aab36336 Salmonella
37	96	12.4	19	3	AAB36328	Aab36328 Salmonella
38	96	12.4	2309	4	ABR66232	ABR66232 Drosophila
39	95.5	12.3	252	8	ADBR3865	ADBR3865 Chemokine
40	95	12.2	597	4	AAU08231	AAU08231 Polypept1
41	94.5	12.2	738	2	AAW6163	AAW6163 New DNA s
42	92.5	11.9	151	3	AAB36344	Aab36344 Escherich
43	92.5	11.9	251	5	ABP45119	ABP45119 Human Bly
44	91	11.7	1397	7	ADD42761	ADD42761 Chlamydia
45	91	11.7	1751	5	ABG91039	ABG91039 Chlamydia

ALIGNMENTS

RESULT 1
AAB36351
ID AAB36351 standard; protein; 151 AA.
XX
AC AAB36351;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#6 amino acid sequence SEQ ID NO:22.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbria; epitope;
KW vaccine; immune response; immunogen.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
XX
N-PSDB; AAC64627.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 137; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (1) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SFP17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, GsgA and AgfA-homologue fimbria subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native

CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fibrillar presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fibrin protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fibrillar subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fibrillae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

XX Sequence 151 AA;

Query Match 100.0%; Score 776; DB 3; Length 151;
 Best Local Similarity 100.0%; Pred. No. 9.5e-68;
 Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLKVAAFAAIVSGSALAGVPPWGGGNNHNGGNSGPDSTLSTIYQGSANAALALQ 60
 DB 1 MKLKVAAFAAIVSGSALAGVPPWGGGNNHNGGNSGPDSTLSTIYQGSANAALALQ 60

QY 61 SDARKSETTITOSGNGADVGAGADNYDQVTRVYTHEMAHADOMNAKNSDITVGOYGG 120
 DB 61 SDARKSETTITOSGNGADVGAGADNYDQVTRVYTHEMAHADOMNAKNSDITVGOYGG 120

QY 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151
 DB 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151

RESULT 2
 AAB36354
 ID AAB36354 standard; protein; 151 AA.
 XX AAB36354;
 AC
 XX 26-FEB-2001 (first entry)
 DT
 XX Agfa::PT3#9 amino acid sequence SEQ ID NO:28.
 DE
 XX *Salmonella*; agfa; chromosomal gene replacement; fibrin; epitope;
 KM vaccine; immune response; immunogen.
 XX
 OS *Salmonella enteritidis*.
 OS *Escherichia coli*.
 OS Synthetic.
 XX
 PN MO20060102-A2.
 PD 12-OCT-2000.
 PF 05-APR-2000; 2000WO-CA000356.
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay WW;
 DR MPI; 2000-672631/65.
 DR N-PSDB; AAC64630.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 138; 139pp; English.

XX The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fibrillae (SEF17/7AF) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Enterobacteriaceae* for the production of fibrillae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fibrin subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fibrillar presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fibrin protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fibrillar subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fibrillae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

XX Sequence 151 AA;

Query Match 91.8%; Score 712; DB 3; Length 151;
 Best Local Similarity 91.1%; Pred. No. 1.7e-61;
 Matches 144; Conservative 0; Mismatches 0; Indels 14; Gaps 2;

QY 1 MKLKVAAFAAIVSGSALAGVPPWGGGNNHNGGNSGPDSTLSTIYQGSANAALALQ 60
 DB 1 MKLKVAAFAAIVSGSALAGVPPWGGGNNHNGGNSGPDSTLSTIYQGSANAALALQ 60

QY 61 SDARKSETTITOSGNGADVGAGADNYDQVTRVYTHEMAHA-----DDMAKNSDI 113
 DB 61 SDARKSETTITOSGNGAD-----YDQVTRVYTHEMAHAFFNNATIDMAKNSDI 113

QY 114 TVGOYGGNNAALVNOTASDSSVMVRQVGFNNATANQY 151
 DB 114 TVGOYGGNNAALVNOTASDSSVMVRQVGFNNATANQY 151

RESULT 3
 AAR74625
 ID AAR74625 standard; protein; 151 AA.
 XX AAR74625;
 AC
 XX 25-MAR-2003 (revised)
 DT 26-JUN-1995 (first entry)
 XX
 DE Agfa sequence.
 XX
 KM *Salmonella*; Agfa; vaccine.
 XX
 OS *Salmonella*.
 OS WO9425598-A2.
 PN 10-NOV-1994.
 PD 26-APR-1994; 94MO-IB000207.
 PF 26-APR-1993; 93US-00054452.
 PR (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
 XX (KING/) KING J.
 XX

PI Kay WW, Collinson SK, Clouthier SC, Doran JL;
 XX WPI; 1994-358275/44.
 DR N-PSDB; AA087467.
 XX
 PT Eliciting an immune response to *Salmonella* - using attenuated *Salmonella*
 PT strains, vector constructs, or compans. contg. fimbrial type proteins.
 XX
 XX Disclosure; Fig 7B; 95pp; English.
 XX
 CC The *Salmonella* Agfa protein and DNA are used in vaccine and genetic
 CC immunization compositions, respectively, to elicit an immune response to
 CC *Salmonella* in animals (e.g. food producing animals) and humans. (Updated
 CC on 25-MAR-2003 to correct PN field.)
 XX
 XX Sequence 151 AA;
 SQ
 Query Match 87.9%; Score 682; DB 2; Length 151;
 Best Local Similarity 90.7%; Pred. No. 1.4e-58;
 Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;
 QY 1 MLLKVAAPAAIVSGSALAGVVPWGGGNNHNGGNSGPPSTLSIYOGSANAALALQ 60
 DB 1 MLLKVAAPAAIVSGSALAGVVPWGGGNNHNGGNSGPPSTLSIYOGSANAALALQ 60
 QY 61 SPARKSETTITGSGNGADVGAGADNDYDQVTRVVTTHMAHADQNNAKNSDITVGOYGG 120
 DB 61 SPARKSETTITGSGNGADVGAGADNSTITELTONGFRNNATIDQNNAKNSDITVGOYGG 120
 QY 121 NNAALVNOTASDSSVMVRQVGFNNATANY 151
 DB 121 NNAALVNOTASDSSVMVRQVGFNNATANY 151
 RESULT 4
 AAB36341
 ID AAB36341 standard; protein; 151 AA.
 XX
 AC AAB36341;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE *Salmonella* enteritidis Agfa amino acid sequence SEQ ID NO:5.
 XX
 KW *Salmonella*; agfa; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS *Salmonella* enteritidis.
 OS
 PN WO200060102-A2.
 PN
 PD 12-OCT-2000.
 PD
 PF 05-APR-2000; 2000WO-CA000356.
 PF
 PR 05-APR-1999; 99US-0127888P.
 PR
 PA (UYVI-) UNIV VICTORIA.
 PA
 PI White AF, Doran JL, Collinson SK, Kay WW;
 XX
 DR WPI; 2000-672631/65.
 DR
 DR N-PSDB; AAC64617.
 DR
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 XX Disclosure; Page 135; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:

CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, Cega and Agfa-homologue fimbria subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 XX Sequence 151 AA;
 SQ
 Query Match 87.9%; Score 682; DB 3; Length 151;
 Best Local Similarity 90.7%; Pred. No. 1.4e-58;
 Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;
 QY 1 MLLKVAAPAAIVSGSALAGVVPWGGGNNHNGGNSGPPSTLSIYOGSANAALALQ 60
 DB 1 MLLKVAAPAAIVSGSALAGVVPWGGGNNHNGGNSGPPSTLSIYOGSANAALALQ 60
 QY 61 SPARKSETTITGSGNGADVGAGADNDYDQVTRVVTTHMAHADQNNAKNSDITVGOYGG 120
 DB 61 SPARKSETTITGSGNGADVGAGADNSTITELTONGFRNNATIDQNNAKNSDITVGOYGG 120
 QY 121 NNAALVNOTASDSSVMVRQVGFNNATANY 151
 DB 121 NNAALVNOTASDSSVMVRQVGFNNATANY 151
 RESULT 5
 AAW23570
 ID AAW23570 standard; protein; 151 AA.
 XX
 AC AAW23570;
 XX
 DT 25-MAR-2003 (revised)
 DT
 DT 29-SEP-1997 (first entry)
 XX
 DE *Salmonella* enteritidis 27655-3b agfa.
 XX
 KW Enteropathogenic bacteria; enterobacteria; *S. enteritidis*; antibody.
 KW
 OS *Salmonella* enteritidis.
 OS
 PN US5635617-A.
 PN
 PD 03-JUN-1997.
 PD
 PF 26-APR-1994; 94US-00233788.
 PF
 PR 26-APR-1993; 93US-00054452.
 PR
 PA (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
 PA
 PI Collinson SK, Kay WW, Doran JL;

```

XX WPI: 1997-309886/28.
DR N-PSDB; AAT74142.
PT Isolated Salmonella gene agfa - used for diagnosis of Salmonella or
XX enteropathogenic bacteria of the Enterobacteria family.
XX Example 2; Fig 7; 85pp; English.
XX The present sequence represents agfa encoded by the full agfa gene
CC derived from Salmonella enteritidis 27655-3b. The nucleic acid can be
CC used to provide diagnostic assays for Salmonella and/or enteropathogenic
CC bacteria of the family Enterobacteria. It can also be used to provide
CC proteins and antibodies which can be used for assays. The nucleic acid
CC sequence can be used to provide probes or primers which can specifically
CC hybridise to nucleic acid molecules from greater than 99% of Salmonella
CC strains that are pathogenic to warm-blooded animals relative to nucleic
CC acid molecules from virtually all other microbial organisms. (Updated on
CC 25-MAR-2003 to correct PF field.)
XX
SQ Sequence 151 AA;
Query Match 87.2%; Score 677; DB 2; Length 151;
Best Local Similarity 90.1%; Pred. No. 4,4e-58;
Matches 136; Conservative 2; Mismatches 13; Indels 0; Gaps 0;
QY 1 MKLTKVAFPAIVSGSALAGVPMQGGGNGHNGSGSPDSTLSTIYQGSANALALQ 60
DB 1 MKLTKVAFPAIVSGSALAGVPMQGGGNGHNGSGSPDSTLSTIYQGSANALALQ 60
QY 61 SDARKSETTTTQSGYGNGADVGGADNYDQLVTRVVTHEMAHADQWNAKNSDITVGGYGG 120
DB 61 SDARKSETTTTQSGYGNGADVGGADNYDQLVTRVVTHEMAHADQWNAKNSDITVGGYGG 120
QY 121 NNAALVNOTASDSSVMRVQGFNNATANOY 151
DB 121 NNAALVNOTASDSSVMRVQGFNNATANOY 151
RESULT 6
AAB36355
ID AAB36355 standard; protein, 151 AA.
AC AAB36355;
XX
XX 26-FEB-2001 (first entry)
XX
XX Agfa::PT3#10 amino acid sequence SEQ ID NO:30.
XX
XX Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
XX vaccine; immune response; immunogen.
XX Salmonella enteritidis.
XX Escherichia coli.
XX Synthetic.
XX
XX WO200060102-A2.
XX
XX 12-OCT-2000.
XX
XX 05-APR-2000; 2000WO-CA000356.
XX
XX 05-APR-1999; 99US-0127888P.
XX
XX (UYVI-) UNIV VICTORIA.
XX
XX White AP, Doran JL, Collison SK, Kay WM;
XX
XX WPI: 2000-672631/65.
XX
XX N-PSDB; AAC64631.
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
XX PT which encodes foreign epitope or antigen, expresses recombinant Agfa

```

```

PT protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 139; 139pp; English.
XX
XX The present invention describes a recombinant agfa gene (1) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/7AF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
XX Sequence 151 AA;
Query Match 84.7%; Score 657; DB 3; Length 151;
Best Local Similarity 81.9%; Pred. No. 4e-56;
Matches 136; Conservative 0; Mismatches 13; Indels 30; Gaps 2;
QY 1 MKLTKVAFPAIVSGSALAGVPMQGGGNGHNGSGSPDSTLSTIYQGSANALALQ 60
DB 1 MKLTKVAFPAIVSGSALAGVPMQGGGNGHNGSGSPDSTLSTIYQGSANALALQ 60
QY 61 SDARKSETTTTQSGYGNGADVGGADN-----YDQLVTRVVTHEMAHADQ 105
DB 61 SDARKSETTTTQSGYGNGADVGGADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHA-- 118
QY 106 WNAKNSDITVGGYGGNNAALVNOTASDSSVMRVQGFNNATANOY 151
DB 106 WNAKNSDITVGGYGGNNAALVNOTASDSSVMRVQGFNNATANOY 151
QY 119 -----GNNALVNOTASDSSVMRVQGFNNATANOY 151
DB 119 -----GNNALVNOTASDSSVMRVQGFNNATANOY 151
RESULT 7
AAB36350
ID AAB36350 standard; protein, 151 AA.
AC AAB36350;
XX
XX 26-FEB-2001 (first entry)
XX
XX Agfa::PT3#5 amino acid sequence SEQ ID NO:20.
XX
XX Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
XX vaccine; immune response; immunogen.
XX Salmonella enteritidis.
XX Escherichia coli.
XX Synthetic.
XX
XX WO200060102-A2.
XX
XX 12-OCT-2000.
XX
XX 05-APR-2000; 2000WO-CA000356.
XX

```

PR 05-APR-1999; 99US-0127888P.
 XX (UYVI-) UNIV VICTORIA.
 PA White AP, Doran JL, Collison SK, Kay WW;
 XX WPI: 2000-672631/65.
 XX N-PSDB; AAC64626.
 DR
 XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 137, 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/7AF) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
 CC Agfa, Caga and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;
 Query Match 79.8%; Score 619; DB 3; Length 151;
 Best Local Similarity 74.6%; Pred. No. 2e-52;
 Matches 129; Conservative 0; Mismatches 0; Indels 44; Gaps 2;
 QY 1 MLLKLVAAFAIIVSGSALAGVVPWGGGNNHNGSGSPDSTLSIYOGSANAALAQ 60
 DB 1 MLLKLVAAFAIIVSGSALAGVVPWGGGNNHNGSGSPDSTLSIYOGSANAALAQ 60
 QY 61 SPARKSETTITOSGYNGADVGGADNVTOLVTRVVTTHMAH----- 103
 DB 61 SPARK-----YQVLVTRVVTTHMAHAGGADNSTTELTONGFR 98
 QY 104 -----DQWNAKNSDITVGOYGNNAAIVNQTASDSSVWVRVYGFNNATANQY 151
 DB 99 NNATIDQWNAKNSDITVGOYGNNAAIVNQTASDSSVWVRVYGFNNATANQY 151
 RESULT 8
 AAB36352
 ID AAB36352 standard; protein; 151 AA.
 XX AAB36352;
 AC
 XX 26-FEB-2001 (first entry)
 DT
 XX Agfa::PT3#7 amino acid sequence SEQ ID NO:24.
 DE
 XX *Salmonella*; agfa; chromosomal gene replacement; fimbrial; epitope;
 KM vaccine; immune response; immunogen.

XX *Salmonella enteritidis*.
 OS *Escherichia coli*.
 OS Synthetic.
 XX
 PN MO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 XX (UYVI-) UNIV VICTORIA.
 PA White AP, Doran JL, Collison SK, Kay WW;
 XX WPI: 2000-672631/65.
 DR N-PSDB; AAC64628.
 DR
 XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 138, 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/7AF) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
 CC Agfa, Caga and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;
 Query Match 78.5%; Score 609; DB 3; Length 151;
 Best Local Similarity 73.6%; Pred. No. 1.9e-51;
 Matches 128; Conservative 0; Mismatches 0; Indels 46; Gaps 2;
 QY 1 MLLKLVAAFAIIVSGSALAGVVPWGGGNNHNGSGSPDSTLSIYOGSANAALAQ 60
 DB 1 MLLKLVAAFAIIVSGSALAGVVPWGGGNNHNGSGSPDSTLSIYOGSANAALAQ 60
 QY 61 SPARKSETTITOSGYNGADVGGADNVTOLVTRVVT-----NYDOLVTRVVT 97
 DB 61 SPARKSETTITOSGYNGADVGGADNVTOLVTRVVT----- 120
 QY 98 HEMAHADQWNAKNSDITVGOYGNNAAIVNQTASDSSVWVRVYGFNNATANQY 151
 DB 121 HEMAH-----NQTASDSSVWVRVYGFNNATANQY 151
 RESULT 9

AAB36346
 ID AAB36346 standard; protein; 151 AA.
 AC AAB36346;
 DT 26-FEB-2001 (first entry)
 DE Agfa::PT#1 amino acid sequence SEQ ID NO:12.
 XX
 XX Salmonella; agfA; chromosomal gene replacement; fimbria; epitope;
 KM vaccine; immune response; immunogen.
 XX
 XX Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 XX WO200060102-A2.
 PD 12-OCT-2000.
 XX
 XX 05-APR-2000; 2000WO-CA000356.
 PF
 XX 05-APR-1999; 99US-0127888P.
 PR
 XX (UYVI-) UNIV VICTORIA.
 PA
 XX White AP, Doran JL, Collison SK, Kay MW;
 PI WPI; 2000-672631/65.
 DR N-PSDB; AAC64623.
 DR
 XX
 XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure: Page 135; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SFF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 XX Sequence 151 AA;
 SQ
 Query Match 77.8%; Score 604; DB 3; Length 151;
 Best Local Similarity 80.8%; Pred. No 5, 9e-51;
 Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;
 QY 1 MLKLKVAAPAAIVVSGSALAGVPPWGGGNNHNGGNSGPDSTLSTIYGYGSANALALQ 60
 DB 1 MLKLKVAAPAAIVVSGSALAGVPPWGGGNNHNGGNSGPDSTLSTIYGYGSANALALQ 60

QY 61 SDARKSETTTTGGGNGADVGGADNDYDQVTRVVTTHMADQNNANSDITVGOYGS 120
 DB 61 SDARKSETTTTGGGNGADVGGADNDYDQVTRVVTTHMADQNNANSDITVGOYGS 120
 QY 121 NNAALVNGTASDSSVWVRQGFENNAATANQY 151
 DB 121 NNAALVNGTASDSSVWVRQGFENNAATANQY 151
 RESULT 10
 AAB36347
 ID AAB36347 standard; protein; 151 AA.
 AC AAB36347;
 DT 26-FEB-2001 (first entry)
 DE Agfa::PT#2 amino acid sequence SEQ ID NO:14.
 XX
 XX Salmonella; agfA; chromosomal gene replacement; fimbria; epitope;
 KM vaccine; immune response; immunogen.
 XX
 XX Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 XX WO200060102-A2.
 PD 12-OCT-2000.
 XX
 XX 05-APR-2000; 2000WO-CA000356.
 PF
 XX 05-APR-1999; 99US-0127888P.
 PR
 XX (UYVI-) UNIV VICTORIA.
 PA
 XX White AP, Doran JL, Collison SK, Kay MW;
 PI WPI; 2000-672631/65.
 DR N-PSDB; AAC64623.
 DR
 XX
 XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure: Page 136; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SFF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in

CC the exemplification of the present invention
XX Sequence 151 AA;
SQ

Query Match 77.6%; Score 602; DB 3; Length 151;
Best Local Similarity 81.5%; Pred. No. 9.3e-51;
Matches 133; Conservative 5; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAAPFAIVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYQGSANALALQ 60
DB 1 MKLLKVAAPFAIVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYQGSANALALQ 60

QY 61 SDAKSETTTTOSGTCNGADVGGADNYDQLVTRVVTHEMAHADQNAKNSDITTVQYGG 120
DB 61 SDAKSETTTTOSGTCNGADVGGADNSTIETLQNGFRNNATIDQNAKNSDITTVQYQ 120

QY 121 NNAALVNOTASDSVWVROVFGNNATANQY 151
DB 121 LVTRVVTHEMAHAGVNGADVGGADNSTIETLQNGFRNNATIDQNAKNSDITTVQYQ 151

RESULT 11
AAB36353
ID AAB36353 standard; protein; 151 AA.
XX
AC AAB36353;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#8 amino acid sequence SEQ ID NO:26.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbria; epitope;
KM vaccine; immune response; immunogen.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UUYI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WM;
XX
DR WPI; 2000-672631/65.
XX
DR N-PSDB; AAC64629.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 138; 139pp; English.
XX
XX The present invention describes a recombinant agfA gene (1) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/7AP) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CagA and AgfA-homologue fimbria subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or

CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention

QY Sequence 151 AA;
SQ

Query Match 77.4%; Score 601; DB 3; Length 151;
Best Local Similarity 81.5%; Pred. No. 1.2e-50;
Matches 123; Conservative 4; Mismatches 24; Indels 0; Gaps 0;

QY 1 MKLLKVAAPFAIVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYQGSANALALQ 60
DB 1 MKLLKVAAPFAIVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYQGSANALALQ 60

QY 61 SDAKSETTTTOSGTCNGADVGGADNYDQLVTRVVTHEMAHADQNAKNSDITTVQYGG 120
DB 61 LVTRVVTHEMAHAGVNGADVGGADNSTIETLQNGFRNNATIDQNAKNSDITTVQYQ 120

QY 121 NNAALVNOTASDSVWVROVFGNNATANQY 151
DB 121 NNAALVNOTASDSVWVROVFGNNATANQY 151

RESULT 12
AAB36349
ID AAB36349 standard; protein; 151 AA.
XX
AC AAB36349;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#4 amino acid sequence SEQ ID NO:18.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbria; epitope;
KM vaccine; immune response; immunogen.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UUYI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WM;
XX
DR WPI; 2000-672631/65.
XX
DR N-PSDB; AAC64625.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 136; 139pp; English.
XX
XX The present invention describes a recombinant agfA gene (1) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:

(1) use of thin aggregative fimbriae (SEB17/TAf) nucleation depended assembly system of strains of *Salmonella*, *Escherichia coli* and *Enterobacteriaceae* for the production of fimbriae comprising recombinant Agfa, CsgA and AgfA-homologue fimbrial subunits, respectively; (2) directing recombination of a recombinant gene into the chromosome of the homologous species; (3) directing recombination of a recombinant gene back into the chromosome of the homologous species, replacing the native copy of that gene; and (4) eliciting an immune response in an animal, comprising separating an amino acid polymer comprising a recombinant Agfa protein containing a replacement segment or segments of foreign amino acid sequence or sequences grown on a *Salmonella*, *E. coli* or *Enterobacteriaceae* host cell, from the host cell and introducing the polymer into the animal in conjunction with a carrier or diluent. (1) is useful for the expression of recombinant Agfa protein which is useful for eliciting an immune response in an animal. In a fimbrial presentation system the heterologous antigens are presented in high numbers (up to 500,000 copies/cell), the hybrid fimbrial protein possesses both the immunogenicity and adhesion properties relevant for an efficient live vaccine, the carrier fimbrial subunit proteins are usually strong immunogens, which may be important for directing an immune response against the inserted epitope, and hybrid fimbriae are easy and inexpensive to purify in large amount. The present sequence is given in the exemplification of the present invention

Sequence 151 AA;

Query Match 77.3%; Score 600; DB 3; Length 151;
Best Local Similarity 81.5%; Pred. No. 1.5e-50;
Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLKVAAFAIVVSGSALAGVVPQWGGGNNHGGNSGPDSTLSTIYQGSANALALQ 60
DB 1 MKLKVAAFAIVVSGSALAGVVPQWGGGNNHGGNSGPDSTLSTIYQGSANALALQ 60
QY 61 SDARKSETTITQSGYGNAGADVGGADNVTDLVTRVYTHEMAHADQNAKNSDITVGYGG 120
DB 61 SDARKSETTITQSGYGNAGADVGGADNVTDLVTRVYTHEMAHADQNAKNSDITVGYGG 120
QY 121 NNAALVNOTASDSSVWVRQVFGNNAATANQY 151
DB 121 NNAALVNOTASDSSVWVRQVFGNNAATANQY 151

RESULT 13

AAB36348 ID AAB36348 standard; protein; 151 AA.

AC AAB36348;

DT 26-FEB-2001 (first entry)

DE Agfa::PT3#3 amino acid sequence SEQ ID NO:16.

KM *Salmonella*; agfa; chromosomal gene replacement; fimbrial; epitope;

KW vaccine; immune response; immunogen.

OS *Salmonella enteritidis*.

OS *Escherichia coli*.

OS synthetic.

PN WO200060102-A2.

PD 12-OCT-2000.

PF 05-APR-2000; 2000WO-CA000356.

PR 05-APR-1999; 99US-0127888P.

XX (UYVT-) UNIV VICTORIA.

PA White AP, Doran JL, Collison SK, Kay WM;

PI WPI; 2000-672631/65.

DR

DR N-PSDB; AAC64624.
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
PS Disclosure; Page 136; 139pp; English.

The present invention describes a recombinant agfa gene (1) where a segment of the gene has been replaced by a segment of a foreign DNA sequence which encodes a foreign epitope or antigen. Also described are: (1) use of thin aggregative fimbriae (SEB17/TAf) nucleation depended assembly system of strains of *Salmonella*, *Escherichia coli* and *Enterobacteriaceae* for the production of fimbriae comprising recombinant Agfa, CsgA and AgfA-homologue fimbrial subunits, respectively; (2) directing recombination of a recombinant gene into the chromosome of the homologous species; (3) directing recombination of a recombinant gene back into the chromosome of the homologous species, replacing the native copy of that gene; and (4) eliciting an immune response in an animal, comprising separating an amino acid polymer comprising a recombinant Agfa protein containing a replacement segment or segments of foreign amino acid sequence or sequences grown on a *Salmonella*, *E. coli* or *Enterobacteriaceae* host cell, from the host cell and introducing the polymer into the animal in conjunction with a carrier or diluent. (1) is useful for the expression of recombinant Agfa protein which is useful for eliciting an immune response in an animal. In a fimbrial presentation system the heterologous antigens are presented in high numbers (up to 500,000 copies/cell), the hybrid fimbrial protein possesses both the immunogenicity and adhesion properties relevant for an efficient live vaccine, the carrier fimbrial subunit proteins are usually strong immunogens, which may be important for directing an immune response against the inserted epitope, and hybrid fimbriae are easy and inexpensive to purify in large amount. The present sequence is given in the exemplification of the present invention

Sequence 151 AA;

Query Match 73.1%; Score 567; DB 3; Length 151;
Best Local Similarity 80.8%; Pred. No. 2.4e-47;
Matches 122; Conservative 5; Mismatches 24; Indels 0; Gaps 0;

QY 1 MKLKVAAFAIVVSGSALAGVVPQWGGGNNHGGNSGPDSTLSTIYQGSANALALQ 60
DB 1 MKLKVAAFAIVVSGSALAGVVPQWGGGNNHGGNSGPDSTLSTIYQGSANALALQ 60
QY 61 SDARKSETTITQSGYGNAGADVGGADNVTDLVTRVYTHEMAHADQNAKNSDITVGYGG 120
DB 61 SDARKSETTITQSGYGNAGADVGGADNVTDLVTRVYTHEMAHADQNAKNSDITVGYGG 120
QY 121 NNAALVNOTASDSSVWVRQVFGNNAATANQY 151
DB 121 NNAALVNOTASDSSVWVRQVFGNNAATANQY 151

RESULT 14

AAB36343 ID AAB36343 standard; protein; 151 AA.

AC AAB36343;

DT 26-FEB-2001 (first entry)

DE *Escherichia coli* CsgA amino acid sequence SEQ ID NO:7.

XX *Salmonella*; agfa; chromosomal gene replacement; fimbrial; epitope;

KW vaccine; immune response; immunogen.

OS *Escherichia coli*.

OS synthetic.

PN WO200060102-A2.

PD 12-OCT-2000.

XX

PF 05-APR-2000; 2000MO-CA000356.
XX
XX 05-APR-1999; 99US-0127888P.
XX
XX (UUYI-) UNIV VICTORIA.
PA
XX
PI White AP, Doran JL, Collison SK, Kay MW;
XX WPI; 2000-672631/65.
DR N-PSDB; AAC64619.
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
XX protein useful for eliciting immune response in animal.
PS Disclosure; Page 135; 139p; English.
XX
XX The present invention describes a recombinant agfa gene (1) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/YAF) nucleation depended
CC assembly system of strains of *Salmonella*, *Escherichia coli* and
CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
CC Agfa, CagA and AgfA-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
CC *Enterobacteriaceae* host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 151 AA;
Query Match 67.1%; Score 521; DB 3; Length 151;
Best Local Similarity 69.5%; Pred. No. 7.5e-43; Indels 0; Gaps 0;
Matches 105; Conservative 17; Mismatches 29;
QY 1 MLLKVAAPAAIVVSGSALAGVVPQWGGGNNHNGSGSPDSTLSIYOGSANAALALQ 60
DB 1 MLLKVAAPAAIVVSGSALAGVVPQWGGGNNHNGSGSPDSTLSIYOGSANAALALQ 60
QY 61 SPARKSETTITGSGYNGADVGQSDNDYLVTRVVTHEMAHADQNNAKNSDITVQYGG 120
DB 61 TPARNSDLITITGSGGNGADVGQSDSDSIDLTQRFGNSATLDQNNKNSSEMTVQFGG 120
QY 121 NNAALVNOTASDSVYVROVFGNNATANQY 151
DB 121 GNGAAVDQTAHSNSVAVTVQVFGNNATANQY 151

RESULT 15

ID ABR82651 standard; protein; 151 AA.

AC ABR82651;

DT 04-DEC-2003 (first entry)

DE E. coli CagA subunit 15 kDa protein.
XX

KW Plasma protein; immune response; antibacterial; vaccine; gene therapy.
XX
XX *Escherichia coli*.
XX
XX WO2003064446-A2.
XX
XX 07-AUG-2003.
XX
XX 30-JAN-2003; 2003WO-EP000943.
XX
XX 31-JAN-2002; 2002GB-00002275.
XX
XX (HANS-) HANSA MEDICAL RES AB.
XX
XX Bjoerck L, Olsen A, Wikstroem M, Herwald H;
XX WPI; 2003-646136/61.
DR N-PSDB; ACF36153.
XX
XX New isolated peptide capable of binding a mammalian plasma protein,
PT useful in the manufacture of a medicament for the prevention and/or
PT treatment of a bacterial infection, such as *Escherichia coli*, *Salmonella*
PT or *Shigella* infections.
XX
XX Disclosure; Page 41-42; 42p; English.
PS
XX
XX The invention relates to an isolated peptide capable of binding a
CC mammalian plasma protein or of generating an immune response in a mammal
CC selected from sequences shown in ABR82647, ABR82648-49. The peptide or
CC antibody is useful for treating a bacterial infection in a human or
CC animal or in the manufacture of a medicament for the prophylactic
CC treatment of a bacterial infection, such as *Escherichia coli*, *Salmonella*
CC or *Shigella* infection. The peptide that is immobilized on a solid support
CC is also useful as a reagent for determining the ability of a plasma
CC protein to bind to bacteria. The present sequence represents an *E. coli*
CC 15 kDa protein
XX
SQ Sequence 151 AA;
Query Match 66.5%; Score 516; DB 7; Length 151;
Best Local Similarity 68.9%; Pred. No. 2.3e-42;
Matches 104; Conservative 17; Mismatches 30; Indels 0; Gaps 0;
QY 1 MLLKVAAPAAIVVSGSALAGVVPQWGGGNNHNGSGSPDSTLSIYOGSANAALALQ 60
DB 1 MLLKVEAIAIAIVFSGSALAGVVPQWGGGNNHNGSGSPNSBINIYOGGNSALALQ 60
QY 61 SPARKSETTITGSGYNGADVGQSDNDYLVTRVVTHEMAHADQNNAKNSDITVQYGG 120
DB 61 TPARNSDLITITGSGGNGADVGQSDSDSIDLTQRFGNSATLDQNNKNSSEMTVQFGG 120
QY 121 NNAALVNOTASDSVYVROVFGNNATANQY 151
DB 121 GNGAAVDQTAHSNSVAVTVQVFGNNATANQY 151

Search completed: March 11, 2004, 18:33:44

Job time : 51.9 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:25:54 ; Search time 12.4 Seconds
(without alignments)
628.671 Million cell updates/sec

Title: US-09-543-407-22

Perfect score: 776
Sequence: 1 MLLKVAAPAAIIVSGSALA.....DSVMVROYFGNNATANY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/2/1aa/5A-COMB.pep:*
2: /cgn2_6/prodata/2/1aa/5B-COMB.pep:*
3: /cgn2_6/prodata/2/1aa/6A-COMB.pep:*
4: /cgn2_6/prodata/2/1aa/6B-COMB.pep:*
5: /cgn2_6/prodata/2/1aa/6C-COMB.pep:*
6: /cgn2_6/prodata/2/1aa/6D-COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	677	87.2	151	US-08-233-788A-59	Sequence 59, App1
2	497	64.0	120	US-08-233-788A-57	Sequence 57, App1
3	94.5	12.2	738	US-08-864-038A-3	Sequence 3, App1
4	87	11.2	892	US-09-336-447A-5	Sequence 5, App1
5	86.5	11.1	238	US-09-495-880A-42	Sequence 42, App1
6	86.5	11.1	585	US-09-620-412C-337	Sequence 337, App
7	86.5	11.1	585	US-09-598-419-337	Sequence 337, App
8	86.5	11.1	1752	US-09-556-877-180	Sequence 180, App
9	86.5	11.1	1752	US-09-620-412C-180	Sequence 180, App
10	86.5	11.1	1752	US-09-598-419-180	Sequence 180, App
11	85.5	11.0	589	US-09-489-039A-7849	Sequence 7849, Ap
12	82	10.6	975	US-09-328-352-4764	Sequence 4764, Ap
13	80	10.3	461	US-09-352-981A-24717	Sequence 24717, A
14	80	10.3	812	US-09-252-991A-30710	Sequence 30710, A
15	80	10.3	1034	US-09-252-991A-26658	Sequence 26658, A
16	79	10.2	943	US-09-056-556-204	Sequence 204, App
17	79	10.2	943	US-09-072-596-199	Sequence 199, App
18	79	10.2	943	US-09-477-135A-131	Sequence 131, App
19	79	10.2	943	US-09-072-967-204	Sequence 204, App
20	79	10.2	2315	US-09-543-681A-5434	Sequence 5434, App
21	78.5	10.1	209	US-09-125-613-42	Sequence 42, App1
22	78.5	10.1	278	US-09-260-283-2	Sequence 2, App1
23	78.5	10.1	309	US-09-252-991A-22266	Sequence 22266, A
24	78.5	10.1	745	US-09-336-115C-6	Sequence 6, App1
25	78	10.1	392	US-08-387-942C-23	Sequence 23, App1
26	78	10.1	553	US-08-387-942C-2	Sequence 2, App1
27	78	10.1	645	US-09-919-172-41	Sequence 41, App1

28	78	10.1	878	US-09-540-236-3401	Sequence 3401, Ap
29	78	10.1	1216	US-09-134-000C-5130	Sequence 5130, Ap
30	77.5	10.0	212	US-09-125-619-32	Sequence 32, App1
31	77.5	10.0	437	US-08-737-716-2	Sequence 2, App1
32	77.5	10.0	673	US-09-196-387-8	Sequence 8, App1
33	77.5	10.0	673	US-09-841-835-8	Sequence 8, App1
34	77.5	10.0	702	US-09-252-991A-22119	Sequence 22119, A
35	77.5	10.0	941	US-09-336-447A-9	Sequence 9, App1
36	77.5	10.0	949	US-09-196-387-10	Sequence 10, App1
37	77.5	10.0	949	US-09-841-835-10	Sequence 10, App1
38	77.5	10.0	1327	US-09-196-387-2	Sequence 2, App1
39	77.5	10.0	1327	US-09-841-835-2	Sequence 2, App1
40	77.5	10.0	1327	US-09-972-115A-8	Sequence 8, App1
41	77.5	10.0	1690	US-09-595-684B-39	Sequence 39, App1
42	77	9.9	266	US-09-495-880A-26	Sequence 26, App1
43	77	9.9	339	US-09-252-991A-32096	Sequence 32096, A
44	77	9.9	873	US-09-336-447A-13	Sequence 13, App1
45	77	9.9	1139	US-08-537-210A-4	Sequence 4, App1

ALIGNMENTS

RESULT 1
US-08-233-788A-59
Sequence 59, Application US/08233788A
Patent No. 5635617
GENERAL INFORMATION:
APPLICANT: Doran, James L.
APPLICANT: Kay, William W.
APPLICANT: Collinson, Karen S.
APPLICANT: Clouthier, Sharon C.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
NUMBER OF INVENTION: OF SALMONELLA
NUMBER OF SEQUENCES: 61
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed and Berry
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: U.S.A.
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233,788A
FILING DATE: 26-APR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: King, Joshua
REGISTRATION NUMBER: 35,570
REFERENCE/DOCKET NUMBER: 920043.403C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEDANDERY
INFORMATION FOR SEQ ID NO: 59:
SEQUENCE CHARACTERISTICS:
LENGTH: 151 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-233-788A-59

Query Match 87.2%; Score 677; DB 1; Length 151;
Best Local Similarity 90.1%; Pred. No. 3.3e-59;
Matches 136; Conservative 2; Mismatches 13; Indels 0; Gaps 0;

QY 1 MLLKVAAPAAIIVSGSALAGVFPWGGGNNNGGSSGPGSTLSIYQGSNAAALIAO 60
|||||

Db 1 MLKLKAAFAAIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANAALALQ 60
QY 61 SPARKSETTITGSGNGADVGQADNDQVTRVVTHEMAHADQMAKNSDITVGYG 120
Db 61 SPARKSETTITGSGNGADVGQADNDSTLTQNGFRNNATIDQMAKNSDITVGYG 120
QY 121 NNAALVNOTASDSVWVRQVGFNNATANQY 151
Db 121 NNPALVNOTASDSVWVRQVGFNNATANQY 151

RESULT 2
US-08-233-788A-57
; Sequence 57, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, Karen S.
; APPLICANT: Clouthier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; TITLE OF INVENTION: OF SALMONELLA
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent'n Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233.788A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 682-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERY
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 120 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-233-788A-57

Query Match 64.0%; Score 497; DB 1; Length 120;
Best Local Similarity 87.5%; Pred. No. 1.2e-41;
Matches 98; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

QY 22 VVPQWGGGNNHNGGNSGPDSTLSIYQGSANAALALQDARKSETTITGSGYGVADY 81
Db 1 VVPQWGGGNNHNGGNSGPDSTLSIYQGSANAALALQDARKSETTITGSGYGVADY 60
QY 82 GQGANNDQVTRVVTHEMAHADQMAKNSDITVGYGANNALVNOTASDS 133
Db 61 GQGANNDSTLTQNGFRNNATIDQMAKNSDITVGYGANNALVNOTASDS 112

RESULT 3
US-08-864-038A-3
; Sequence 3, Application US/08864038A
; Patent No. 6001592
; GENERAL INFORMATION:

; APPLICANT: Kunio NAKASHIMA et al.
; TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE CDNA, VECTOR
; TITLE OF INVENTION: CONTAINING SAID CDNA, HOST CELLS TRANSFORMED WITH SAID
; TITLE OF INVENTION: VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING
; TITLE OF INVENTION: SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBODY
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: 812-5 Hiranoo
; STREET: Ieshinden
; CITY: Tsu-city
; STATE: Mie-prefecture
; COUNTRY: JAPAN
; ZIP: 514-01
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Microsoft Windows 95
; SOFTWARE: Word Perfect 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/864,038A
; FILING DATE: May 28, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-184459
; FILING DATE: 15-July-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: C. Bruce Handburg
; REGISTRATION NUMBER: 22,389
; REFERENCE/DOCKET NUMBER: F-5610
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)986-2340
; TELEFAX: (212)953-7733
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 738
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Pinctada fucata
; CELL TYPE: mantle epithelial cell
; FEATURE:
; NAME/KEY: peptide
; LOCATION: from 1 to 738
; IDENTIFICATION METHOD: E (by experiment)
; US-08-864-038A-3

Query Match 12.2%; Score 94.5; DB 3; Length 738;
Best Local Similarity 27.6%; Pred. No. 0.48;
Matches 43; Conservative 14; Mismatches 58; Indels 41; Gaps 6;

QY 3 LKLKAAFAAIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANAALALQSD 62
Db 419 LKSSASASASASASAG-----GGGGGNGGNGGNGG-----GGGAGALAA----- 460

QY 63 ARKSETTITGSGNGADVGQADNDQVTRVVTHEMAHADQMAKNSDITVGYGVGN 121
Db 461 -----ALAAAGAGGGLGGGGG-----ALAAALAAAGAGGGGCGTGGAGGGG 507

QY 122 NNAALVNOTASDS-----VWVRQVGFNNATAN 148
Db 508 SNAALVNOTASDS-----VWVRQVGFNNATAN 148

RESULT 4
US-09-336-447A-5
; Sequence 5, Application US/09336447A
; Patent No. 6310190
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: ABEI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL

```

:
: APPLICANT: FISKE, MICHAEL J.
: APPLICANT: FREDENBURG, ROSS A.
: TITLE OF INVENTION: USPA1 AND USPA2 ANTIGENS OF MORAXELLA CATARRHALIS
: FILE REFERENCE: AMCT:024
: CURRENT APPLICATION NUMBER: US/09/336,447A
: CURRENT FILING DATE: 1999-06-21
: NUMBER OF SEQ. ID NOS.: 98
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 5
:
: LENGTH: 892
:
: TYPE: PR1
:
: ORGANISM: Moraxella catarrhalis
:
US-09-336-447A-5

```

```

Query Match      11.2%; Score 87; DB 4; Length 892;
Best Local Similarity 26.8%; Pred. No. 3.4;
Matches 38; Conservative 15; Mismatches 57; Indels 32; Gaps 7;

QY 28 GGGNHN-----GGGNSGSPSTLSIYOYGSAANLALQSDARKSETTITOSGY--- 75
    |||::||
Db 89 GGCGPYNAKGNYSITYVGGSSNTAKGEKSTICGGDPTN-----DANGTYSTIGCYYSRA 141
    |||::||

QY 76 -GNADVGCGGDNDTDLVTRYVTHEMAADGMNAKSIDITYGYG--GNNAAVL----N 127
    |||::||
Db 142 IGDSTTIGCG--YYNQATGEKSTVAGGRNNOATGNNSTVAGGSYNQATGNNSTVAGGSNN 199
    |||::||

QY 128 QTASDSIVMVRQVFGNNATAN 149
    |||::||
Db 200 QATGESF---AAGVENKANNN 218
    |||::||

```

```

RESULT 5
US-09-495-880A-42
; Sequence 42: Application US/09495880A
; Patent No. 6667150
; GENERAL INFORMATION:
; APPLICANT: RUDERT, FRITZ
; APPLICANT: GE. LIMIING
; APPLICANT: ILAG, VIC
; TITLE OF INVENTION: NOVEL METHOD AND PHAGE FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES ENCODING MEMBERS OF A MULTIMERIC
; TITLE OF INVENTION: (POLY)PEPTIDE COMPLEX
; FILE REFERENCE: MORPHO/9
; CURRENT APPLICATION NUMBER: US/09/495,880A
; CURRENT FILING DATE: 2000-02-01
; PRIOR APPLICATION NUMBER: PCT/EP98/04836
; PRIOR FILING DATE: 1998-08-03
; PRIOR APPLICATION NUMBER: EP 97 11 3319.4
; PRIOR FILING DATE: 1997-08-01
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 42
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Artificial Sequence
FEATURES:
; OTHER INFORMATION: Description of Artificial Sequence: ompA-FLAG-peptide3-
; OTHER INFORMATION: gene iits encoded by phage vector fpep3_1B-IR3seq (circular
US-09-495-880A-42

```

Query Match	11.1%	Score 86.5	DB 4	Length 238
Best Local Similarity	24.3%	Pred. No. 0.69		
Matches	43	Conservative	24	Mismatches 63
				Indels 47
				Gaps 8
Qy	5	KVAAPAIIVVGSALAG-----	-VYPOMGGGNNH-NGGSGNSGPDSTLSI	47
Db	3	KTALAIIVALLAGFATVADKYDVCITYAHAIYLVAKGGGGSEFNAGGGSGG-----		55
Qy	48	YQYGSANAALALQSDARKSEFTTITQSGYGNAGDVGQAGDNTDQLVTRVVTHEMAHADQWN		107
Db	56	---GSGGSGEGGSGEGGSGEGGSGEGC-GSGGSGSGSDPVEKMANNAKAMTEADE-N		110
Qy	108	AKNSDI-----TYGQYG-----GNNALVNTQA-----SDSSVMRYQVFGNNA		146

```

Db
111 ALGSDAKGKLDVSATDYGAALDIDPGIDVSGLANGATGATPGAGNSQAAGVGDGDN5 167

RESULT 6
US-09-620-412C-337
Sequence 337, Application US/09620412C
Patent No. 6448234
GENERAL INFORMATION:
APPLICANT: Steven P. Fling
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
FILE REFERENCE: 210121.469C7
CURRENT APPLICATION NUMBER: US/09/620,412C
CURRENT FILING DATE: 2000-07-20
NUMBER OF SEQ ID NOS: 363
SOFTWARE: FastSeq for Windows Version 3.0/4.0
SEQ ID NO 337
LENGTH: 585
TYPE: PRT
ORGANISM: Chlamydia trachomatis
US-09-620-412C-337

```

Query Match	11.1%	Score 86.5	DB 4	Length 585
Best Local Similarity	31.6%	Pred No. 2.2		
Matches	43	Conservative 12	Mismatches 56	Indels 23
			Gaps	7
QY	4	LKVAFPAALIVSGSALAGVVPQWGGCGN--HNGGNSGSPDS---TLSIYQGSANAAL	57	
Db	206	LKQKQASAG--NADMAVSSPQSGGATVPSDGSQSSGSDPSRSETVPYAKG-----	25	
QY	58	ALQSDARKSETTIT--QSGYGNAGDVQCGADNTQLYTRVYVTHMAHADPMNAKNSDIT	115	
Db	258	GLVTDNRLSLITITGIIEIIRANNAKRDVGGGA-----YKGTLLTCENSHRLOFLKNSDPKQ	312	
QY	115	VGO-VGGANNAALVNOT	129	
Db	313	GGGIYGEDNITLISNLT	328	

```

RESULT 7
US-09-598-419-337
; Sequence 337, Application US/09598419
; Patent No. 6565856
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Scholler, John
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
; FILE REFERENCE: 210121.469C6
; CURRENT APPLICATION NUMBER: US/09/598,419
; CURRENT FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 357
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 337
; LENGTH: 585
; TYPE: PR1
; ORGANISM: Chlamydia trachomatis
; US-09-598-419-337

```

	Query Match	11.1%;	Score 86.5;	DB 4,	Length 585;	
	Best Local Similarity	31.6%;	Pred. No. 2.2;			
	Matches	43;	Conservative 12;	Mismatches 58;	Indels 23;	Gaps 7
QY	4 LKVAAPAAIVVSGSALAGVPMGGGNN--HNGGSNSGSPDS---TLSTIYQGSANAAL	57				
Dd	206 LKQAASAC---NADMASSSPQSSGATTVSDSGSSSSGSDSPSETVPYIAKG-----	25				
QY	58 ALQSDARKSETTIT---QSRYGNQADVQGDADNYDQLVTRVYTHEMAHADQMNAKNSDIT	114				
Dd	258 GLVTIDNLSTINITGTIEIANNKARDVVGGA-----YKGLTTCENSHRLDFLNKSSDKQ	314				
QY	115 VQG-YGQNPAALVNQT	129				

Db 313 GGGIYGEDNITLSNLT 328

RESULT 8
US-09-556-877-180

; Sequence 180, Application US/09556877

; Patent No. 6432916

; GENERAL INFORMATION:

; APPLICANT: Probst, Peter

; APPLICANT: Bhatia, Ajay

; APPLICANT: Skeiky, Yasir

; APPLICANT: Fling, Steve

; APPLICANT: Maisonneuve, Jeff

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND

; FILE REFERENCE: 210121.469C6

; CURRENT FILING DATE: US/09/556.877

; NUMBER OF SEQ ID NOS: 305

; SOFTWARE: FastSeq for Windows Version 3.0/4.0

; SEQ ID NO 180

; LENGTH: 1752

; TYPE: PRT

; ORGANISM: Chlamydia

US-09-556-877-180

Query Match 11.1%; Score 86.5; DB 4; Length 1752;
Best Local Similarity 31.6%; Pred. No. 9;

Matches 43; Conservative 12; Mismatches 58; Indels 23; Gaps 7;

Db 388 LKAQASAG--NADWASSSPGSGATTVSDSDSSGSDTSETVPTAKG----- 439

QY 4 LKVAFAAIVVSGSALAGVPPWGGGNN--HNGGNSGSPDS---TLSIYGSANAAL 57

Db 58 ALQSDARKSETTIT--OSGYNGADVGGADNYDQVTRVVTHEMAHADQNAKNSDIT 114

QY 440 GLYTDKNLSTNITGIIIEIANNKATDVGGG-----YKGTITCENSHRLOFLKNSSDKQ 494

Db 115 VGO-YGNNALVNOT 129

QY 495 GGGIYGEDNITLSNLT 510

RESULT 9

US-09-620-412C-180

; Sequence 180, Application US/09620412C

; Patent No. 6448234

; GENERAL INFORMATION:

; APPLICANT: Steven P. Fling

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND

; FILE REFERENCE: 210121.469C7

; CURRENT FILING DATE: US/09/620.412C

; NUMBER OF SEQ ID NOS: 363

; SOFTWARE: FastSeq for Windows Version 3.0/4.0

; SEQ ID NO 180

; LENGTH: 1752

; TYPE: PRT

; ORGANISM: Chlamydia

US-09-620-412C-180

Query Match 11.1%; Score 86.5; DB 4; Length 1752;
Best Local Similarity 31.6%; Pred. No. 9;

Matches 43; Conservative 12; Mismatches 58; Indels 23; Gaps 7;

QY 4 LKVAFAAIVVSGSALAGVPPWGGGNN--HNGGNSGSPDS---TLSIYGSANAAL 57

Db 388 LKAQASAG--NADWASSSPGSGATTVSDSDSSGSDTSETVPTAKG----- 439

QY 58 ALQSDARKSETTIT--OSGYNGADVGGADNYDQVTRVVTHEMAHADQNAKNSDIT 114

Db 440 GLYTDKNLSTNITGIIIEIANNKATDVGGG-----YKGTITCENSHRLOFLKNSSDKQ 494

QY 115 VGO-YGNNALVNOT 129

Db 495 GGGIYGEDNITLSNLT 510

RESULT 10

US-09-598-419-180

; Sequence 180, Application US/09598419

; Patent No. 6565856

; GENERAL INFORMATION:

; APPLICANT: Skeiky, Yasir A.W.

; APPLICANT: Scholler, John

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT AND

; FILE REFERENCE: 210121.469C6

; CURRENT FILING DATE: US/09/598.419

; NUMBER OF SEQ ID NOS: 357

; SOFTWARE: FastSeq for Windows Version 3.0/4.0

; SEQ ID NO 180

; LENGTH: 1752

; TYPE: PRT

; ORGANISM: Chlamydia

US-09-598-419-180

Query Match 11.1%; Score 86.5; DB 4; Length 1752;
Best Local Similarity 31.6%; Pred. No. 9;

Matches 43; Conservative 12; Mismatches 58; Indels 23; Gaps 7;

Db 388 LKAQASAG--NADWASSSPGSGATTVSDSDSSGSDTSETVPTAKG----- 439

QY 4 LKVAFAAIVVSGSALAGVPPWGGGNN--HNGGNSGSPDS---TLSIYGSANAAL 57

Db 58 ALQSDARKSETTIT--OSGYNGADVGGADNYDQVTRVVTHEMAHADQNAKNSDIT 114

QY 440 GLYTDKNLSTNITGIIIEIANNKATDVGGG-----YKGTITCENSHRLOFLKNSSDKQ 494

Db 115 VGO-YGNNALVNOT 129

QY 495 GGGIYGEDNITLSNLT 510

RESULT 11

US-09-489-039A-7849

; Sequence 7849, Application US/09489039A

; Patent No. 6610836

; GENERAL INFORMATION:

; APPLICANT: Gary Bretton et. al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA

; FILE REFERENCE: 2709.2004001

; CURRENT FILING DATE: US/09/489.039A

; PRIOR APPLICATION NUMBER: 2000-01-27

; PRIOR FILING DATE: 1999-01-29

; NUMBER OF SEQ ID NOS: 14342

; SEQ ID NO 7849

; LENGTH: 589

; TYPE: PRT

; ORGANISM: Klebsiella pneumoniae

US-09-489-039A-7849

Query Match 11.0%; Score 85.5; DB 4; Length 589;
Best Local Similarity 23.2%; Pred. No. 2.8;

Matches 32; Conservative 17; Mismatches 68; Indels 21; Gaps 3;

QY 29 GGNHNGNSGSPSTLSIYOGS-----ANAALALQSDARKSETTITQSGYNGAD 80

Db 259 GGLDRNGANANGQDITGIIAFDITLTERIEINGRLDNYHTKYSATACGSSGRGAI 318

QY 81 VGGADNYDQVTRVVTHEMAHADQNA-----KNSDITVGOYGNNALVNOTASDS 133

Db 319 ACPGGSTGSPVTTVTATSGNLVNWKAGALYRLTEQANVY-----NYALISQGPGRGS 372
QY 134 SVMVRQVGFNNATANQY 151
Db 373 SFRLAASGSGNSANRTDF 390

RESULT 12

US-09-328-352-4764
; Sequence 4764, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNI FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4764
; LENGTH: 975
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-4764

Query Match 10.6%; Score 82; DB 4; Length 975;
Best Local Similarity 24.1%; Pred. No. 12;

Matches 38; Conservative 15; Mismatches 47; Indels 58; Gaps 7;

QY 15 GSGALAGVVPQWGGGNNHNGG-GNSSGPDSTLSIYQY-----SANA 55
Db 300 AGNGIA-----SGNGEHNYGIGNGGDDVDITAPITGVNLISGNSFTLIGNSSSVNT 353
QY 56 ALALQSDAKSEITI-----TQSGYG-----NGADVGGADNYDQLVTRVYTHEM 100
Db 354 APTTSTNTVNDNDITDINGSGGTGSGSGSGSGGDLNGAASGNGEHNYG----- 402
QY 101 AHADQWNAKNSDIT-----VGQYGNNAALVNOTASDS 133
Db 403 --IGNGGDDVDITSPITGTGFNSGNSFLIGNSSSSS 438

RESULT 13

US-09-252-991A-24717
; Sequence 24717, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24717
; LENGTH: 461
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24717

Query Match 10.3%; Score 80; DB 4; Length 461;
Best Local Similarity 25.6%; Pred. No. 7;

Matches 43; Conservative 20; Mismatches 67; Indels 38; Gaps 8;

QY 6 VAAFAAIVVSG-SALAGVVPQWGGGNNHNGGSSGPDSTLSIYQGSANA---ALALQSD 61
Db 300 VAALEPVANSGPAGPAGTAPAGGAA---CGKSPAGLGRILARRSPSSSATPPAATYTF 356

QY 62 DARKSETTITQSG-YANGA-----DVGGADNYDQLVT-----RVYTHEMAHADQMN- 107
Db 357 AARAPWPAITPERGSGGTGAPDRPRRTAGTGEQGLVVKQFAPPAYRLVYQGLIEGDSWNG 416
QY 108 -----AKNSDITVQYGGNNALVNOTASDSVWVRQVGFNNA 146
Db 417 DVFLRIDMAELQGFDFWGIADGNQAVVASQVISET-----VGNHA 456

RESULT 14

US-09-252-991A-30710
; Sequence 30710, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30710
; LENGTH: 812
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30710

Query Match 10.3%; Score 80; DB 4; Length 812;
Best Local Similarity 26.7%; Pred. No. 15;

Matches 31; Conservative 10; Mismatches 39; Indels 36; Gaps 5;

QY 51 GSNALALQSDARKSE-----TTTQSGYGNADVGGADNYDQLVTRVYTHEMAHAD 104
Db 550 GLINIGVTRDSRYSERYIWSRSTPSQGLGWNLYGGGASRYQD-----AD 597
QY 105 -QWNAKNSDITVQYGG--GN-----NALVNOTASDSVWVRQVGF 142
Db 598 LTRWQNVOLQGGGLVGETGNTYRMADLSGSLYWMQNAVNASRINRINDAPLVSTKGY 653

RESULT 15

US-09-252-991A-26658
; Sequence 26658, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 26658
; LENGTH: 1034
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-26658

Query Match 10.3%; Score 80; DB 4; Length 1034;
Best Local Similarity 26.7%; Pred. No. 20;

Matches 47; Conservative 19; Mismatches 62; Indels 48; Gaps 10;

QY 14 VGSALAGVVPQWGGGNNHNGGSSGPDSTLSIYQGS-ANALALQSDAR-- 64
Db 748 VSDSSASQGV-EAGGAGNTGLVGLSGGEIRFSQASGVSYSKGLATGLIGKAEQNGM 806

OY 65 ----XSETTITQSGYGADYQGAGADNYDOLV-----TRY----- 95
| : : | | | | |
Db 807 LGNLIKASGSVTDOG---GADLGGLVGNNSQSAIETAEATGKVSQGSNSRVGGLIGHNLOG 863
OY 96 -VTHEMAHADQWNAKNSDITVQYGGNNAALVNOTASD--SSVMVRQYG--FGNNA 146
| : : | | | | |
Db 864 SVAAHAISRQDVSGGFNS-LVGGLVGHNGGELVNVVDASGRVSAALASASVGLVGSNA 918

Search completed: March 11, 2004, 18:44:52
Job time : 13.4 secs

Db 149 SVIODGKNNVFSIKQNTGNTSTVNQIGEMNAVYRQIGIAETDASTGNLPTGANNV 208
 QY 124 ALVNOTASDSSVWVQVGFSGNN 145
 Db 209 ASITONSAGLNTAVAVGGGNS 230

RESULT 2 US-09-793-306-146

; Sequence 146, Application US/09793306
 ; Patent No. US20020098200A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Campos-Neto, Antonio
 ; APPLICANT: Skeiky, Yasir
 ; APPLICANT: Ovendale, Pamela
 ; APPLICANT: Jen, Shyian
 ; APPLICANT: Lodes, Michael
 ; APPLICANT: Corixa Corporation
 ; TITLE OF INVENTION: Compounds and Methods for Diagnosis and Immunotherapy
 ; TITLE OF INVENTION: of Tuberculosis
 ; FILE REFERENCE: 014058-008740US
 ; CURRENT APPLICATION NUMBER: US/09/793,306
 ; PRIOR FILING DATE: 2001-02-26
 ; PRIOR APPLICATION NUMBER: US 60/185,037
 ; PRIOR FILING DATE: 2000-02-25
 ; PRIOR APPLICATION NUMBER: US 60/223,828
 ; PRIOR FILING DATE: 2000-08-08
 ; NUMBER OF SEQ ID NOS: 164
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 146
 ; LENGTH: 597
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: mTTC#3-His
 ; US-09-793-306-146

Query Match 12.2%; Score 95; DB 9; Length 597;
 Best Local Similarity 27.4%; Pred. No. 0.36;
 Matches 34; Conservative 14; Mismatches 50; Indels 26; Gaps 4;

QY 26 WGGGNNHGGNNSGPDSTLSIYQGSANALALQSDARKSETTTTOSGYNAGADVQGA 85
 Db 358 FGSGNNNIGFNSG-NNNVGFNSGNN-----FGFNAGDINTGF 398
 QY 86 DNYDQLVTVVTHEMAHADQNNAKNSDITVGOYGANNAALVNOTASDSSVWVROVGFSGNN 145
 Db 399 GNAGDVTNFGNAGFNNIGMAGNEDMGVNGSGSFVVGYN--AGNOS-----VGFQNA 451
 QY 146 ATAN 149
 Db 452 GTLN 455

RESULT 3 US-09-880-748-1130

; Sequence 1130, Application US/09880748
 ; Publication No. US2003005937A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Ruben et al.
 ; TITLE OF INVENTION: Antibodies that Immunospecifically Bind Blys
 ; FILE REFERENCE: P5523
 ; CURRENT APPLICATION NUMBER: US/09/880,748
 ; CURRENT FILING DATE: 2001-06-15
 ; PRIOR APPLICATION NUMBER: 60/212,210
 ; PRIOR FILING DATE: 2000-06-15
 ; PRIOR APPLICATION NUMBER: 60/240,816
 ; PRIOR FILING DATE: 2000-10-17
 ; PRIOR APPLICATION NUMBER: 60/276,248
 ; PRIOR FILING DATE: 2001-03-16
 ; PRIOR APPLICATION NUMBER: 60/277,379
 ; PRIOR FILING DATE: 2001-03-21
 ; PRIOR APPLICATION NUMBER: 60/293,499

; PRIOR FILING DATE: 2001-05-25
 ; NUMBER OF SEQ ID NOS: 3239
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 1130
 ; LENGTH: 251
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-09-880-748-1130

Query Match 11.9%; Score 92.5; DB 10; Length 251;
 Best Local Similarity 27.7%; Pred. No. 0.22;
 Matches 33; Conservative 14; Mismatches 51; Indels 21; Gaps 4;

QY 18 ALAGVVPQW-----CGGNHGGNNSGPDSTLSIYQGSANALALQSDARKSE- 67
 Db 106 ATTGALDMMKGTGLTVTVSSGGGGGGGGGG-----GSAQAVLTQPSASGTRG 155
 QY 68 TTTTOSGYNGADVGGADN-YDQLVTVVTHEMAHADQNNAKNSDITVGOYGANNAAL 125
 Db 156 QRTVMCGSSSSNIGSNTVNTWYQLPGAAPKLIYRSDRSSGVPRFGSGKSGTSASL 214

RESULT 4 US-09-841-132-445

; Sequence 445, Application US/09841132
 ; Patent No. US20020061848A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Bhatia, Ajay
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Probst, Peter
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
 ; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
 ; FILE REFERENCE: 210121.469C8
 ; CURRENT APPLICATION NUMBER: US/09/841,132
 ; PRIOR FILING DATE: 2001-04-23
 ; NUMBER OF SEQ ID NOS: 599
 ; SOFTWARE: FastSeq for Windows Version 3.0/4.0
 ; SEQ ID NO 445
 ; LENGTH: 1751
 ; TYPE: PRT
 ; ORGANISM: Chlamydia trachomatis serovar D
 ; US-09-841-132-445

Query Match 11.7%; Score 91; DB 9; Length 1751;
 Best Local Similarity 33.1%; Pred. No. 3.5;
 Matches 40; Conservative 10; Mismatches 59; Indels 12; Gaps 5;

QY 15 SGSAALAGVVPWGGGNN--HNGGNSGPDSTLSIYQGSANALALQSDARKSETTTT- 71
 Db 394 NADAMASSSPGSGGATVSNSGDSSGSDSTSEVPATAKG-GLYTDKVLSTINITG 452
 QY 72 --OSGYNGADVGGADNDQLVTVVTHEMAHADQNNAKNSDITVGO-YGGNNALVNO 128
 Db 453 IIEIANAKATDVGGG-----YKGTITCENSHRLQFLKNSDKGGGIYGBDNTLSNL 507
 QY 129 T 129
 Db 508 T 508

RESULT 5 US-09-841-132-594

; Sequence 594, Application US/09841132
 ; Patent No. US20020061848A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Bhatia, Ajay
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Probst, Peter
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
 ; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
 ; FILE REFERENCE: 210121.469C8
 ; CURRENT APPLICATION NUMBER: US/09/841,132
 ; PRIOR FILING DATE: 2001-04-23

; NUMBER OF SEQ ID NOS: 599
 ; SOFTWARE: FastSeq for Windows Version 3.0/4.0
 ; SEQ ID NO 594
 ; LENGTH: 1751
 ; TYPE: PRF
 ; ORGANISM: C. Trachomatis D serovar
 US-09-841-132-594

Query Match 11.7%; Score 91; DB 9; Length 1751;
 Best Local Similarity 33.1%; Pred. No. 3.5;
 Matches 40; Conservative 10; Mismatches 59; Indels 12; Gaps 5;

QY 15 GSGALGVTPWGGG--HNGGNSGPDSTLSTYQGSANAALALQSDAKSETTIT- 71
 DB 394 NADAWASSPSSGATVTSNSGSSGSDSTSETPATAKG-GLYTDKNLSTINIG 452
 QY 72 --OSGNGADYGGGADNTDQVTRVTHEMAHADQMNKNSDITVGO-YGGNNALVNG 128
 DB 453 IIEIANKATDVGGG-----YKGTLTGNSHRLQPLKNSSDKGGGIGEDNITLSNL 507
 QY 129 T 129
 DB 508 T 508

RESULT 6

US-09-880-748-1122
 ; Sequence 1122; Application US/09880748
 ; Publication No. US2003005937A1

; GENERAL INFORMATION:
 ; APPLICANT: Ruben et al.
 ; TITLE OF INVENTION: Antibodies that Immunospecifically Bind Blys
 ; FILE REFERENCE: PF523
 ; CURRENT APPLICATION NUMBER: US/09/880,748
 ; PRIOR FILING DATE: 2001-06-15
 ; PRIOR APPLICATION NUMBER: 60/212,210
 ; PRIOR FILING DATE: 2000-06-15
 ; PRIOR APPLICATION NUMBER: 60/240,816
 ; PRIOR FILING DATE: 2000-10-17
 ; PRIOR APPLICATION NUMBER: 60/276,248
 ; PRIOR FILING DATE: 2001-03-16
 ; PRIOR APPLICATION NUMBER: 60/277,379
 ; PRIOR FILING DATE: 2001-03-21
 ; PRIOR APPLICATION NUMBER: 60/293,499
 ; PRIOR FILING DATE: 2001-05-25
 ; NUMBER OF SEQ ID NOS: 3239
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 1122
 ; LENGTH: 251
 ; TYPE: PRF
 ; ORGANISM: Homo sapiens
 US-09-880-748-1122

Query Match 11.5%; Score 89.5; DB 10; Length 251;
 Best Local Similarity 28.9%; Pred. No. 0.43;
 Matches 35; Conservative 15; Mismatches 46; Indels 25; Gaps 5;

QY 18 ALAGVTPWGG-----GGGHNHNGSGSDSTLSTYQGSANAALALQ---SDAK 65
 DB 106 ATTGALDMWGGTLTVSSGGGGGGGGG-----GSAOSVLTQPPSVSEAR 155
 QY 66 SETTITQSGYNGADYGGGADN-YDQVTRVTHEMAHADQMNKNSDITVGOYGGNNAA 124
 DB 156 QRTVITCS--GSAENVGNNAVMVYQQLPKGPKLLIYYDDLPSGVSDRPSGSGSGTSAS 213
 QY 125 L 125
 DB 214 L 214

RESULT 7
 US-10-369-493-20619
 ; Sequence 20619; Application US/10369493

Publication No. US20030233675A1

GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

FILE REFERENCE: 38-10(52052)B

CURRENT APPLICATION NUMBER: US/10/369,493

PRIOR FILING DATE: 2002-02-21

NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 20619

LENGTH: 486

TYPE: PRF

ORGANISM: Rhodospseudomonas palustris

FEATURE:

NAME/KEY: unsure

LOCATION: (1)..(486)

OTHER INFORMATION: unsure at all Xaa locations

US-10-369-493-20619

Query Match 11.3%; Score 87.5; DB 15; Length 486;
 Best Local Similarity 22.9%; Pred. No. 1.6;
 Matches 33; Conservative 24; Mismatches 52; Indels 35; Gaps 5;

QY 11 AIVS--ALAGVTPWGGGHNHNGGNSGP-----DSTLSTYQGSANAAL 57
 DB 86 SVVAGTVDYRDVIAAGLPGLGKAYALGDSGPAPKDLASDPFSAPEGAADGTV 145
 QY 58 ALQSDARKSETTITQSGYNGADYGGGADNVDQVTRVTHEMAHADQMNKNSDITVGO- 116
 DB 146 ITH-----TAAVGRRGALISG-----NLIASLVDAIRLTREADVNLGM 188
 QY 117 ---QYGGNNALVNGTASDSSVM 136
 DB 189 LPLFVYTGGLMLTLQAGAGASVI 212

RESULT 8

US-09-952-267-5

Sequence 5; Application US/09952267

Publication No. US20030032772A1

GENERAL INFORMATION:

; APPLICANT: HANSEN, ERIC J.
 ; APPLICANT: AEBI, CHRISTOPH
 ; APPLICANT: COPE, LESLIE D.
 ; APPLICANT: MACIVER, ISOBEL
 ; APPLICANT: FISKE, MICHAEL J.
 ; APPLICANT: FREDENBURG, ROSS A.

TITLE OF INVENTION: USPA1 AND USPA2 ANTIGENS OF MORAXELLA CATARRHALIS

FILE REFERENCE: AMCY:024

CURRENT APPLICATION NUMBER: US/09/952,267

PRIOR FILING DATE: 2001-09-12

PRIOR APPLICATION NUMBER: 09/336,447

PRIOR FILING DATE: 1998-06-21

NUMBER OF SEQ ID NOS: 98

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 5

LENGTH: 892

TYPE: PRF

ORGANISM: Moraxella catarrhalis

US-09-952-267-5

Query Match 11.2%; Score 87; DB 10; Length 892;
 Best Local Similarity 26.8%; Pred. No. 3.8;
 Matches 38; Conservative 15; Mismatches 57; Indels 32; Gaps 7;

QY 28 GGGNN-----GGGNSGPDSTLSTYQGSANAALALQSDAKSETTITQSGY--- 75
 |||:|

Db 89 GGGDYNEAKNYSITVGGSSNTAKGEKSTIGGDTN-----DANGTYSITIGGYSRA 141
Qy 76 -GNGADVCGANNYQULVTRVYTHEMAHADOMNANSDITVQYCG--GNNALV-----N 127
Db 142 IGDSSTIGGG--YVQATGEEKSTVAGGRNNQATGNNSITVAGSYNQATGNNSITVAGGSHN 199
Qy 128 QTASDSVWVRQVGFNNATAN 149
Db 200 QATGEGSF---AAGVENKANNAN 218

RESULT 9

US-09-841-132-337
; Sequence 337, Application US/09841132
; Patent No. US20020061848A1
; GENERAL INFORMATION:
; APPLICANT: Bhatia, Ajay
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Probst, Peter
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
; FILE REFERENCE: 210121.469C8
; CURRENT APPLICATION NUMBER: US/09/841.132
; NUMBER OF SEQ ID NOS: 599
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 337
; LENGTH: 585
; TYPE: PRT
; ORGANISM: Chlamydia trachomatis
US-09-841-132-337

Query Match 11.1%; Score 86.5; DB 9; Length 585;
Best Local Similarity 31.6%; Pred. No. 2.5;
Matches 43; Conservative 12; Mismatches 58; Indels 23; Gaps 7;

Qy 4 LKVAAPFAIVSGSALAGVVPWGGGNN--HNGGNSGPDG---TLSITVYOGSANAAL 57
Db 206 LKAQASAG--NADAMASSPQSGGATVVSDDSSGSDPTSETVPTAKG----- 257
Qy 58 ALQSDARKSETTT---QSGYNGADVGGGADNDQULVTRVYTHEMAHADOMNANSDIT 114
Db 258 GLYIDKMLSTNTITGIIETIANNKATDVGGG-----YKGTLCENSHRLQFLKNSDKQ 312
Qy 115 VQG-YGGNNALVNOT 129
Db 313 GGGIYGEDNITLSNLT 328

RESULT 10

US-09-841-132-180
; Sequence 180, Application US/09841132
; Patent No. US20020061848A1
; GENERAL INFORMATION:
; APPLICANT: Bhatia, Ajay
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Probst, Peter
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
; FILE REFERENCE: 210121.469C8
; CURRENT APPLICATION NUMBER: US/09/841.132
; NUMBER OF SEQ ID NOS: 599
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 180
; LENGTH: 1752
; TYPE: PRT
; ORGANISM: Chlamydia
US-09-841-132-180

Query Match 11.1%; Score 86.5; DB 9; Length 1752;
Best Local Similarity 31.6%; Pred. No. 10;
Matches 43; Conservative 12; Mismatches 58; Indels 23; Gaps 7;

Qy 4 LKVAAPFAIVSGSALAGVVPWGGGNN--HNGGNSGPDG---TLSITVYOGSANAAL 57
Db 388 LKAQASAG--NADAMASSPQSGGATVVSDDSSGSDPTSETVPTAKG----- 439
Qy 58 ALQSDARKSETTT---QSGYNGADVGGGADNDQULVTRVYTHEMAHADOMNANSDIT 114
Db 440 GLYIDKMLSTNTITGIIETIANNKATDVGGG-----YKGTLCENSHRLQFLKNSDKQ 494
Qy 115 VQG-YGGNNALVNOT 129
Db 495 GGGIYGEDNITLSNLT 510

RESULT 11

US-09-820-843A-21
; Sequence 21, Application US/09820843A
; Publication No. US2003003963A1
; GENERAL INFORMATION:
; APPLICANT: Council of Scientific and Industrial Research
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTE
; FILE REFERENCE: 063915
; CURRENT APPLICATION NUMBER: US/09/820.843A
; CURRENT FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 21
; LENGTH: 354
; TYPE: PRT
; ORGANISM: M. tuberculosis
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: PPE
; NAME/KEY: misc feature
; OTHER INFORMATION: g11781260
US-09-820-843A-21

Query Match 11.0%; Score 85; DB 10; Length 354;
Best Local Similarity 22.8%; Pred. No. 1.9;
Matches 38; Conservative 13; Mismatches 58; Indels 58; Gaps 6;

Qy 10 AAIIVSGSALAG-VVPWGGGNNHNGG-----NSSGPDSTLSITVYGSAN 54
Db 200 AGDVNSGVGNAGDVNTGNGSGNINTGFPNGELNTGFFSAMTQAGPNS--GFFNAGCN 257
Qy 55 AALALQSDARKSETTTTQSGYNGADVGGGADNDQULVTRVYTHEMAHADOMNANSDIT 114
Db 258 SGFGHNDPAGSGNSGICQNSGFNGSGYVNTS-----T 288
Qy 115 VQYCGN-----NNAALVNOTASDSSVWVRQY---GFGNNATAN 149
Db 289 TSMFGNSGVNTGNGSGFNNAAVNNTGIFVTGVMSSGFFNGCN 335

RESULT 12

US-09-880-748-2098
; Sequence 2098, Application US/09880748
; Publication No. US20030059937A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Antibodies that Immunospecifically Bind Blys
; FILE REFERENCE: PFS23
; CURRENT APPLICATION NUMBER: US/09/880.748
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 60/212,210
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: 60/240,816
; PRIOR FILING DATE: 2000-10-17
; PRIOR APPLICATION NUMBER: 60/276,248
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/277,379
; PRIOR FILING DATE: 2001-03-21

```

; PRIOR APPLICATION NUMBER: 60/293,499
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 3239
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2098
; LENGTH: 253
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-880-748-2098

Query Match
Best Local Similarity 10.8%; Score 84; DB 10; Length 253;
Matches 34; Conservative 16; Mismatches 46; Indels 34; Gaps 7;

QY 7 AAFPAIVSGSALAGVVP-----QWG-----GGNNHNGGNSGPPDSTLSIYOGSANAALALQSDA 53
DB 92 AVFYCVREGGADYDAFYFDWGGCTLVTVSSGGSGSGSGSGG-----GSA 141
QY 54 NAALALQSDARKSE-TTITQSGYGADVGQADN--YDQL---VTRVVTHEMAH----- 102
DB 142 QAVLTQSSSVSGAPGQGVITISCTGSSSNIGDYVHWYQLPETAPEKLIYGNSHRPSGV 201
QY 103 ADQWNAKNSD 112
DB 202 PDRFSGSKSD 211

RESULT 13
US-09-880-748-1153
; Sequence 1153, Application US/09880748
; Publication No. US2003005937A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Antibodies that Immunospecifically Bind Blys
; FILE REFERENCE: PF523
; CURRENT APPLICATION NUMBER: US/09/880,748
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 60/212,210
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: 60/240,816
; PRIOR FILING DATE: 2000-10-17
; PRIOR APPLICATION NUMBER: 60/276,248
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/277,379
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/293,499
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 3239
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1153
; LENGTH: 255
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-880-748-1153

Query Match
Best Local Similarity 10.8%; Score 84; DB 10; Length 255;
Matches 24; Conservative 11; Mismatches 31; Indels 20; Gaps 3;

QY 13 VVSGSALAGVPMWG-----GGNNHNGGNSGPPDSTLSIYOGSANAALALQSDA 63
DB 104 ILTYGWSGAFDQWKGITWTVVSSGGSGSGSGSGG-----GSAQAVLTQPPSSV 153
QY 64 RKSE-TTITQSGYGADVGQADNY 88
DB 154 SGAPGQGVITISCTGSSSNIGAGYDVY 179

RESULT 14
US-09-880-748-1494
; Sequence 1494, Application US/09880748
; Publication No. US2003005937A1
; GENERAL INFORMATION:
```

```

; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Antibodies that Immunospecifically Bind Blys
; FILE REFERENCE: PF523
; CURRENT APPLICATION NUMBER: US/09/880,748
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 60/212,210
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: 60/240,816
; PRIOR FILING DATE: 2000-10-17
; PRIOR APPLICATION NUMBER: 60/276,248
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/277,379
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/293,499
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 3239
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1494
; LENGTH: 257
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-880-748-1494

Query Match
Best Local Similarity 10.8%; Score 83.5; DB 10; Length 257;
Matches 26; Conservative 13; Mismatches 31; Indels 21; Gaps 4;

QY 13 VVSGSALAGVPMWG-----GGNNHNGGNSGPPDSTLSIYOGSANAALALQSDA 63
DB 104 ILTYGWSGAFDQWKGITWTVVSSGGSGSGSGSGG-----GSAQAVLTQPPSSV 153
QY 64 RKSE-TTITQSGYGADVGQAD-NYDQLV 92
DB 154 SGAPGQGVITISCTGSSSNIGAGYEVHWYQLV 184

RESULT 15
US-09-880-748-1199
; Sequence 1199, Application US/09880748
; Publication No. US2003005937A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Antibodies that Immunospecifically Bind Blys
; FILE REFERENCE: PF523
; CURRENT APPLICATION NUMBER: US/09/880,748
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 60/212,210
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: 60/240,816
; PRIOR FILING DATE: 2000-10-17
; PRIOR APPLICATION NUMBER: 60/276,248
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/277,379
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/293,499
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 3239
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1199
; LENGTH: 253
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-880-748-1199

Query Match
Best Local Similarity 10.7%; Score 83; DB 10; Length 253;
Matches 31; Conservative 13; Mismatches 50; Indels 26; Gaps 4;

QY 22 VVPMWG-----GGNNHNGGNSGPPDSTLSIYOGSANAALALQSDARKSB 67
DB 107 VTRARMDVWGGCTLVTVSSGGSGSGSGSGG-----GSAQAVLTQPPSSASGTP 156
QY 68 -TTITQSGYGADVGQADNYDQLVTRVVTHEMAH-DQWNAKNSDITVGGYGGNNAL 125
```

Fri Mar 12 09:41:58 2004

us-09-543-407-22.rapb

Page 6

Db 157 GQRTTISGSGSSNIGSNVINWYQQLPRTAPKLIHTNDQRPBGVDPDRFSGSKSGTSGTL 216

Search completed: March 11, 2004, 19:18:38
Job time : 24.6 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:24:14 ; Search time 10.3 Seconds

(without alignments)
1410.186 Million cell updates/sec

Title: US-09-543-407-22

Perfect score: 776
Sequence: 1 MLKLKVAAPAAIVSGSALA.....DSSVWVQVFGNNATANOY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	682	87.9	151	2 JC6039	fimbria protein ag
2	682	87.9	151	2 AI0635	major curlin chain
3	521	67.1	151	2 S70788	curlin protein csg
4	499.5	64.4	152	2 D90806	curlin major subun
5	499.5	64.4	152	2 H85665	hypothetical prote
6	109	14.0	1748	2 S42136	cnjB protein - Tet
7	109	14.0	2174	2 E95965	hypothetical glyci
8	102	13.1	586	2 T26667	hypothetical prote
9	98	12.6	635	2 E70663	probable PPE prote
10	95	12.2	530	2 E70946	probable PPE prote
11	94	12.1	407	2 T21956	hypothetical prote
12	93.5	12.0	145	2 AD3143	conserved hypotet
13	93.5	12.0	145	2 H98144	hypothetical prote
14	93.5	12.0	645	2 F70825	probable PPE prote
15	93.5	12.0	1567	2 S11572	ice nucleation pro
16	92.5	11.9	151	2 S70787	curlin nucleator p
17	92.5	11.9	151	2 C90806	minor curlin subun
18	92.5	11.9	151	2 G85665	curlin minor chain
19	91.5	11.8	1034	2 JC2143	ice nucleation act
20	91.5	11.8	1258	2 JQ0188	ice nucleation pro
21	91	11.7	1751	2 G71518	hypothetical prote
22	90.5	11.7	151	2 JC6040	fimbria protein ag
23	90.5	11.7	151	2 AH0635	nucleation compone
24	89.5	11.5	1053	2 B70987	probable PPE prote
25	89.5	11.5	1332	2 S07053	ice nucleation pro
26	89	11.5	575	2 S55327	protein kinase agg
27	87	11.2	251	2 D6010	hypothetical expor
28	87	11.2	434	2 E70768	hypothetical glyci
29	87	11.2	573	2 C86266	f3f19.21 protein -

30	87	11.2	590	1 A45621	leishmanolysin (EC
31	87	11.2	599	2 B42049	leishmanolysin (EC
32	87	11.2	599	2 A44951	leishmanolysin (EC
33	87	11.2	602	1 PL0221	leishmanolysin (EC
34	87	11.2	646	1 S19916	leishmanolysin (EC
35	86.5	11.1	262	2 S00275	tail fiber protein
36	86.5	11.1	487	2 C70830	probable PPE prote
37	86.5	11.1	1028	2 A56038	DNA-binding protei
38	86.5	11.1	1213	2 S16356	ovo protein - fru
39	86	11.1	582	2 F70675	probable PPE prote
40	85.5	11.0	940	2 D89723	protein F39D8.1b f
41	85.5	11.0	945	2 T21998	hypothetical prote
42	85.5	11.0	1210	2 A25547	ice nucleation pro
43	85	11.0	354	2 B70663	probable PPE prote
44	85	11.0	1052	2 AF2959	conserved hypotet
45	85	11.0	1341	2 H98323	hypothetical prote

ALIGNMENTS

RESULT 1

JC6039
fimbria protein agfa precursor - Salmonella enteritidis

C/Spect: Salmonella enteritidis

C/Date: 31-Dec-1996 #sequence, revision 31-Dec-1996 #ext_change 08-Oct-1999

C/Accession: JC6039; PC6015; A44898

R/Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Baner, P.A.; Kay, W.W.

J. Bacteriol. 178, 662-667, 1996

A/Title: Salmonella enteritidis agfA operon encoding thin, aggregative fimbriae.

A/Reference number: JC6039; PMID:96146512; PMID:8550497

A/Accession: JC6039

A/Molecule type: DNA

A/Residues: 1-151 <COL>

A/Cross-references: GB:U43280; NID:G1184712; PID:NAC43599.1; PID:G1184714

A/Accession: PC6015

A/Molecule type: protein

A/Residues: 21-52 <CO2>

A/Experimental source: strain 27655-3b

A/Note: the authors translated the codon ACG for residue 44 as Ile

R/Collinson, S.K.; Emdy, L.; Muller, K.H.; Trust, T.J.; Kay, W.W.

J. Bacteriol. 173, 4773-4781, 1991

A/Title: Purification and characterization of thin, aggregative fimbriae from Salmonella

A/Reference number: A44898; PMID:91310586; PMID:1677357

A/Accession: A44898

A/Status: preliminary

A/Molecule type: protein

A/Residues: 21-33 <CO3>

A/Note: sequence extracted from NCBI backbone (NCBIP:45936)

C/Genetics:

A/Genes: agfa

C/Function:

A/Description: major component of thin aggregative fimbriae

A/Note: fimbriae bind to fibronectin, plasminogen, tissue plasminogen activator

C/Keywords: fimbria

F/1-20/Domain: signal sequence #status predicted <SIG>

F/21-151/Product: fimbria protein agfa #status experimental <MAT>

Query Match 87.9%; Score 682; DB 2; Length 151;

Best Local Similarity 90.7%; Pred. No. 1.2e-50;

Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

QY	1	MLKLKVAAPAAIVSGSALAGVPPOMGGGNNHGGSSGPDSTLSIYOGSNNALALQ	60
DB	1	MLKLKVAAPAAIVSGSALAGVPPOMGGGNNHGGSSGPDSTLSIYOGSNNALALQ	60
QY	61	SDARKSETTITSGYNGADVGADNDYDQVTVTVTHMAADQNNASDITVOYGG	120
DB	61	SDARKSETTITSGYNGADVGADNDYDQVTVTVTHMAADQNNASDITVOYGG	120
QY	121	NNALVNTASDSSVWVROVFGNNATANOY 151	
DB	121	NNALVNTASDSSVWVROVFGNNATANOY 151	

Db 121 NNAALVNTASDSSVMVQVGFNNATANQY 151

RESULT 2

AIO635

major curlin chain precursor [imported] - Salmonella enterica subsp. enterica serovar Typhimurium

A/Species: Salmonella enterica subsp. enterica serovar Typhimurium

A/Note: this species has also been called Salmonella typhimurium

C/Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002

C/Accession: AIO635

R:Parthill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Churcher, T.; Conerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A/Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.

A/Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhimurium

A/Reference number: AB0502; MUID:21534947; PMID:11677608

A/Accession: AIO635

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-151 <PAR>

A/Cross-references: GB:AL513382; PIDN:CAD08268.1; PID:gl6502315; GSPDB:GN00176

C/Genetics:

A/Gene: STY1181

Query Match 87.9%; Score 682; DB 2; Length 151;
Best Local Similarity 90.7%; Pred. No. 1.2e-50;
Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

QY 1 MRLKVAFAAIVVSGSALAGVVPQWGGGNNHGGNSGPDSTLTIYOGSANAALALQ 60
DB 1 MRLKVAFAAIVVSGSALAGVVPQWGGGNNHGGNSGPDSTLTIYOGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADYVQGGADNYDQVTRVVTHEMAHADOMNAKNSDITVQYGG 120
DB 61 SDARKSETTITQSGYNGADYVQGGADNYDQVTRVVTHEMAHADOMNAKNSDITVQYGG 120
QY 121 NNAALVNTASDSSVMVQVGFNNATANQY 151
DB 121 NNAALVNTASDSSVMVQVGFNNATANQY 151

RESULT 3

S70788

curlin protein csgA precursor - Escherichia coli (strain K-12)

N/Alternate names: csgA protein; major curlin protein

C/Species: Escherichia coli

C/Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 01-Mar-2002

C/Accession: S70788; G64846; S31202; S34560; S34559

R:Hammar, M.; Arngvist, A.; Bian, Z.; Olsen, A.; Normark, S.

Mol. Microbiol. 18, 661-670, 1995

A/Title: Expression of two csg operons is required for production of fibronectin- and C

A/Reference number: S70783; MUID:96414468; PMID:8817489

A/Accession: S70788

A/Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-151 <HAM>

A/Cross-references: EMBL:X90754; NID:g1147558; PIDN:CAA62282.1; PID:g1147564

A/Experimental source: strain K12, substrain W3110

A/Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co

A.; Rose, D.J.; Mau, B.; Shaio, Y.

Science 271, 1453-1462, 1997

A/Title: The complete genome sequence of Escherichia coli K-12.

A/Reference number: A64720; MUID:97426617; PMID:9278503

A/Accession: G64846

A/Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-151 <BLAT>

A/Cross-references: GB:AE000205; GB:U00096; NID:g1787265; PIDN:AACT4126.1; PID:g1787279;

A/Experimental source: strain K-12, substrain MG1655

R:Olsen, A.; Arngvist, A.; Hammar, M.; Sukupolvi, S.; Normark, S.

Mol. Microbiol. 7, 523-536, 1993

A/Title: The Rpos sigma factor relieves H-NS-mediated transcriptional repression of csgA

A/Reference number: S31202; MUID:93211294; PMID:8459772

A/Accession: S31202

A/Molecule type: DNA

A/Residues: 1-6, 'V', 8-151 <OLS1>

A/Cross-references: EMBL:U04979

A/Accession: S34560

A/Molecule type: protein

A/Residues: 21-42; 44-50 <OLS2>

R:Olsen, A.N.; Arngvist, A.M.

submitted to the EMBL Data Library, October 1992

A/Reference number: S34559

A/Accession: S34559

A/Molecule type: DNA

A/Residues: 1-133, 'RORDSGMTW' <OLS3>

A/Cross-references: EMBL:U04979; NID:g290424; PIDN:AAA23616.1; PID:g290425

A/Experimental source: strain K-12, substrain W3110

C/Genetics:

A/Gene: csgA

A/Map position: 23.15

C/Function:

A/Description: major component of wild-type curli; interaction between CsgA and CsgB tr and H-kinogen; in the absence of CsgB, CsgA can self-assemble into polymers

F:1-20/Domain: signal sequence #status predicted <SIG>

F:21-151/Product: curlin #status experimental <MAT>

Query Match 67.1%; Score 521; DB 2; Length 151;
Best Local Similarity 69.5%; Pred. No. 4.6e-37;
Matches 105; Conservative 17; Mismatches 29; Indels 0; Gaps 0;

QY 1 MRLKVAFAAIVVSGSALAGVVPQWGGGNNHGGNSGPDSTLTIYOGSANAALALQ 60
DB 1 MRLKVAFAAIVVSGSALAGVVPQWGGGNNHGGNSGPDSTLTIYOGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADYVQGGADNYDQVTRVVTHEMAHADOMNAKNSDITVQYGG 120
DB 61 SDARKSETTITQSGYNGADYVQGGADNYDQVTRVVTHEMAHADOMNAKNSDITVQYGG 120
QY 121 NNAALVNTASDSSVMVQVGFNNATANQY 151
DB 121 NNAALVNTASDSSVMVQVGFNNATANQY 151

RESULT 4

D90806

curlin major subunit CsgA [imported] - Escherichia coli (strain O157:H7, substrain RIMD

C/Species: Escherichia coli

C/Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001

C/Accession: D90806

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G

gatawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.

DNA Res. 8, 11-22, 2001

A/Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen

A/Reference number: A99629; MUID:21156231; PMID:11258796

A/Accession: D90806

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-152 <HAY>

A/Cross-references: GB:BA000077; PIDN:BA834843.1; PID:g13360880; GSPDB:GN00154

A/Experimental source: strain O157:H7, substrain RIMD 0509952

C/Genetics:

A/Gene: ECS1420

Query Match 64.4%; Score 499.5; DB 2; Length 152;
Best Local Similarity 67.8%; Pred. No. 3e-35;
Matches 103; Conservative 18; Mismatches 30; Indels 1; Gaps 1;

QY 1 MRLKVAFAAIVVSGSALAGVVPQWGGGNNHGGNSGPDSTLTIYOGSANAALALQ 59
DB 1 MRLKVAFAAIVVSGSALAGVVPQWGGGNNHGGNSGPDSTLTIYOGSANAALALQ 60
QY 60 SDARKSETTITQSGYNGADYVQGGADNYDQVTRVVTHEMAHADOMNAKNSDITVQYGG 119

Db 61 QARARSDLTITTHGGGNGADVGGSDSSIDLTRQGFNSATLTDQNGKSHMTYKQFG 120
 QY 120 GNNAAVLNQTASDSSVWVQVFGNNATANQY 151
 121 GGGGAADVDTASNSTVNVTVQVFGNNATANQY 152

RESULT 5

H85665
 hypothetical protein cega [imported] - Escherichia coli (strain O157:H7, substrain EDL93
 C/Species: Escherichia coli
 C/Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 C/Accession: H85665
 R/Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.U.; Mayhew
 Miller, L.; Grobbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
 Nature 409, 529-533, 2001
 A/Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A/Reference number: A85480; MUID:21074935; PMID:11206551
 A/Accession: H85665
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-152 <STO>
 A/Cross-references: GB:AE005174; NID:G12514574; PIDN:AG55788.1; GSPDB:GND0145; UMGF:Z16
 A/Experimental source: strain O157:H7, substrain EDL933
 C/Genetics:
 A/Genes: cega

Query Match 64.4%; Score 499.5; DB 2; Length 152;
 Best Local Similarity 67.8%; Pred. No. 3e-35;
 Matches 103; Conservative 18; Mismatches 30; Indels 1; Gaps 1;

QY 1 MKLLKTAAPFAIIVSSGALAGVPPW-GGGGNNHGGNSGPDSTLSIYQGSANALAL 59
 1 MKLLKTAAPFAIIVSSGALAGVPPW-GGGGNNHGGNSGPDSTLSIYQGSANALAL 60
 DB 60 QSDARSETTTTGGGNGADVGGADNDTQDLTVRVVTHEMAHADQNAKNSDITVQY 119
 61 QADARRSDLTITTHGGGNGADVGGSDSSIDLTRQGFNSATLTDQNGKSHMTYKQFG 120
 QY 120 GNNAAVLNQTASDSSVWVQVFGNNATANQY 151
 121 GGGGAADVDTASNSTVNVTVQVFGNNATANQY 152

RESULT 6

S42136
 cnB protein - Tetrahymena thermophila
 C/Species: Tetrahymena thermophila
 C/Date: 19-Mar-1997 #sequence_revision 18-Jul-1997 #text_change 07-Dec-1999
 C/Accession: S42136; S42135; S03650
 R/Taylor, F.M.; Martindale, D.W.
 Submitted to the EMBL Data Library, October 1992
 A/Reference number: S42136
 A/Accession: S42136
 A/Molecule type: DNA
 A/Residues: 1-1748 <TAY>
 A/Cross-references: EMBL:L03710; NID:G161751; PID:G161752
 R/Taylor, F.M.; Martindale, D.W.
 Nucleic Acids Res. 21, 4610-4614, 1993
 A/Title: Retroviral-type zinc fingers and glycine-rich repeats in a protein encoded by c
 A/Reference number: S42135; MUID:94051569; PMID:8233798
 A/Accession: S42135
 A/Molecule type: DNA
 A/Residues: 1164-1174,1179-1198,1233-1252,1285-1293,1297-1309,1316-1326,1331-1341,1343-1
 A/Cross-references: EMBL:L03710
 R/Martindale, D.W.; Taylor, F.M.
 Nucleic Acids Res. 16, 2189-2201, 1988
 A/Title: Multiple functions in a conjugation-specific gene from Tetrahymena thermophila.
 A/Reference number: S03650; MUID:88189811; PMID:3357771
 A/Accession: S03650
 A/Molecule type: DNA
 A/Residues: 236-250, 'I', 252-255, 'N', 257-773 <MAR>

A/Cross-references: EMBL:X06462
 C/Genetics:
 A/Genes: cnB
 A/Genetic code: SGC5
 A/Intron: 85/3; 136/1; 157/3; 201/2; 290/2; 327/3; 499/1; 573/2; 607/3; 708/3; 777/3;
 C/Keywords: zinc finger
 F/1164-1450/Region: glycine-rich
 F/1451-1464/Region: zinc finger CCHC motif
 F/1478-1491/Region: zinc finger CCHC motif
 F/1501-1514/Region: zinc finger CCHC motif
 F/1530-1543/Region: zinc finger CCHC motif
 F/1555-1568/Region: zinc finger CCHC motif
 F/1579-1592/Region: zinc finger CCHC motif
 F/1602-1615/Region: zinc finger CCHC motif
 F/1626-1748/Region: glycine-rich

Query Match 14.0%; Score 109; DB 2; Length 1748;
 Best Local Similarity 33.6%; Pred. No. 0.39; 32; Indels 30; Gaps 6;
 Matches 37; Conservative 11; Mismatches 30; Indels 30; Gaps 6;

QY 25 OMGGGNNHGG---GNSGPDSTLSIYQGSANALALQSDARKSETTT---QSGYGN 77
 1640 QFGGGGNSGGGSGWSSGSDNN-----CGSNVOESTTSSGGGSSGSGN 1685
 DB 78 GADVQGA-DNTDQTVRVVTHEMAHADQNAKNSDITVQY--YGGNNA 124
 DB 1686 QFGGGGNSGNDNOQ-----QNTGCGGGSNSQNTNSESWSGNNQA 1729

RESULT 7

E95965
 hypothetical glycine-rich protein [imported] - Sinorhizobium meliloti (strain 1021) mag
 C/Species: Sinorhizobium meliloti
 C/Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
 C/Accession: E95965
 R/Finan, T.M.; Weidner, S.; Wong, K.; Bhramasree, J.; Chain, P.; Vorholter, F.J.; Herna
 Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
 A/Title: The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing end
 A/Reference number: A95842; MUID:21396508; PMID:11481431
 A/Accession: E95965
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-2174 <KUR>
 A/Cross-references: GB:AL591985; PIDN:CAC49389.1; PID:G15140875; GSPDB:GND0167
 A/Experimental source: strain 1021, megaplasmid pSymB
 R/Galbert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
 pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federpiel, N.A.; Fisher, R.F.,
 L.; Hyman, R.W.; Jones, T.
 Science 293, 668-672, 2001
 A/Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kles, E.; Komp, C.; Lelaure,
 heault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Weller, D.H.; Wong, K.; Yeh, K
 A/Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
 A/Reference number: A96039; MUID:21368234; PMID:11474104
 A/Contents: annotation
 C/Genetics:
 A/Genes: Smb21548
 A/Genome: plasmid

Query Match 14.0%; Score 109; DB 2; Length 2174;
 Best Local Similarity 27.0%; Pred. No. 0.49; 52; Indels 36; Gaps 7;
 Matches 40; Conservative 20; Mismatches 36; Indels 36; Gaps 7;

QY 11 AIVSSGALAGVPPQ--WGGGNNHGGNSGPDSTLSIYQGS-----ANAA----- 56
 693 AATAGAGAVGILAGISGGGNN---GSAATGDAFGFSQGGGGGGGYANTANVGR 749
 DB 57 -IALOSDARKSETTTTGGGNGADVGGADNDTQDLTVRVVTHEMAHADQNAKNSDITV 115
 750 GLTLFTTQSGHAAGIYQVS-VGGGGGCGTASYSAGI-----GFTASVAV 793
 QY 116 GQYGGNNA--LVNQTASDSSVWVQV 141
 DB 794 GGTGGNGAGGVSLSLTDSAIRTQGG 821

RESULT 8

T26667
 C/Species: Caenorhabditis elegans
 C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C/Accession: T26667
 R/Mallis, J.
 Submitted to the EMBL Data Library, September 1999
 A/Reference number: Z20252
 A/Accession: T26667
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-586 <WTL>
 A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A/Experimental source: clone Y38E10A
 C/Genetics:
 A/Introns: 105/2; 174/1; 248/1; 372/3; 436/3

Query Match 13.1%; Score 102; DB 2; Length 586;
 Best Local Similarity 28.8%; Pred. No. 0.44;
 Matches 47; Conservative 14; Mismatches 66; Indels 36; Gaps 7;

QY 4 LKVAFAFAIVVSGSALAGVVPQWGGGNNH-----GGNSGSPDSTLSIYOGSANAALAL 59
 DB 437 VRSVGFQAQOFQSGQFAPPIPAGGGGSGCYGAGGSGG-----YGAGGAGGAR 487
 QY 60 QSDARKSETTITGSGYNGAD---VQGADNYDQVTRVVT-----HENAHDQWNAKN 110
 DB 488 NSASNGS-----YSSANEVKS VGFQAQVGSVFAKPSCTGGGVASGSGARSGE 539
 QY 111 SDITVGOYGNNALVNOTAS--DSSVMRVQVFGNNATANOY 151
 DB 540 SGAGGGAGGGKAGAKNNSYSSANEVKSVPFG---AQOY 578

RESULT 9

E70663
 probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
 C/Species: Mycobacterium tuberculosis
 C/Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 15-Sep-2003
 C/Accession: E70663
 R/Cole, S.T.; Brosch, R.; Parthill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Fellwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A/Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A/Reference number: A70500; MUID:98295987; PMID:9634230
 A/Accession: E70663
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-615 <COL>
 A/Cross-references: GB:Z83860; GB:AL123456; NID:93261681; PIDN:CA806165.1; PID:e290763;
 A/Experimental source: strain H37RV
 C/Genetics:
 A/Gene: PPE

Query Match 12.6%; Score 98; DB 2; Length 615;
 Best Local Similarity 22.5%; Pred. No. 1;
 Matches 31; Conservative 16; Mismatches 45; Indels 46; Gaps 5;

QY 15 SGAALAGVVPQWGGGNNHGGNSGPDSTLSIYOGSANAALALQSDARKSETTITGSG 74
 DB 365 SGTGNIG---FGNSGNNTIGFNSG-DGNIGFNSGDN-----TG 401
 QY 75 YGNADVGGADNDQVTRVTHMAHADQWNAKNSDITVGOYGNNALV---NOTAS 131
 DB 402 FGNAGNINTGF-----WNGNLTGFGSANGVIGFDGNSNSG 441
 QY 132 DSSVMRVQVFGNNATAN 149

DB 442 SFNVGFQNTGFGNSGAGN 459

RESULT 10

E70946
 probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
 C/Species: Mycobacterium tuberculosis
 C/Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 15-Sep-2003
 C/Accession: E70946
 R/Cole, S.T.; Brosch, R.; Parthill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Fellwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A/Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A/Reference number: A70500; MUID:98295987; PMID:9634230
 A/Accession: E70946
 A/Molecule type: DNA
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown
 A/Residues: 1-590 <COL>
 A/Cross-references: GB:AL021646; GB:AL123456; NID:93242278; PIDN:CAA16624.1; PID:e12487
 A/Experimental source: strain H37RV
 C/Genetics:
 A/Gene: PPE

Query Match 12.2%; Score 95; DB 2; Length 590;
 Best Local Similarity 27.4%; Pred. No. 1.7;
 Matches 34; Conservative 14; Mismatches 50; Indels 26; Gaps 4;

QY 26 WGGGNNHGGNSGPDSTLSIYOGSANAALALQSDARKSETTITGSGYNGADVQGA 85
 DB 351 FGSGNNNTIGFNSG-NNNVGFNSGNN-----FGFGNAGDINTGF 391
 QY 86 DNYDQVTRVTHMAHADQWNAKNSDITVGOYGNNALVNOTASDSSVMRVQVFGNN 145
 DB 392 GNAAGDNTGFGNAGFPNMGIGNAGNEDGVNGSGSFNVGVN--AGNQ-----VGFNA 444
 QY 146 ATAN 149
 DB 445 GTLN 448

RESULT 11

T21956
 hypothetical protein F38B7.3 - Caenorhabditis elegans
 C/Species: Caenorhabditis elegans
 C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C/Accession: T21956
 R/Lennard, N.
 Submitted to the EMBL Data Library, June 1996
 A/Reference number: Z19493
 A/Accession: T21956
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-407 <WTL>
 A/Cross-references: EMBL:Z74033; PIDN:CAA98477.1; GSPDB:GN00023; CESP:F38B7.3
 A/Experimental source: clone F38B7
 C/Genetics:
 A/Gene: CESP:F38B7.3
 A/Map position: 5
 A/Introns: 12/1; 57/3; 124/2; 163/1; 330/3

Query Match 12.1%; Score 94; DB 2; Length 407;
 Best Local Similarity 25.8%; Pred. No. 1.4;
 Matches 34; Conservative 19; Mismatches 43; Indels 36; Gaps 7;

QY 9 FAIVVSGSALA-GVVPQWGG-----GGNHHGGNSGPDSTLSIYOGSANAALALQ 60
 DB 250 FGSGILAGSLGLGSLSMGHHHSYGGWGGYGGG-----YGMAG---GYV 293
 QY 61 SDARKSETTITGSGYNGADVQGADNDQVTRVTHMAHADQWNAKNSDITVGOYGG 120

Db 294 SD---NDTVTNNYNYGNDNNGNNHSS-SSNALTDAFGNAGDSQQQSD----- 341

QY 121 NNAALVNOTASD 132

Db 342 NNYGNANQDSYD 353

RESULT 12

AD3143

conserved hypothetical protein Atcu4768 [imported] - Agrobacterium tumefaciens (strain C58)

C/Species: Agrobacterium tumefaciens

C/Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002

C/Accession: AD3143

R/Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.

erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavlin, T.; Levy, R.; Li, M.; McClellan

Science 294, 2317-2323, 2001

A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,

ster, E.W.

A/Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A/Reference number: AB2577; PMID:21608550; PMID:11743193

A/Accession: AD3143

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-145 <RUR>

A/Cross-references: GB:AB008669; PIDN:AA145562.1; PID:g17743277; GSPDB:GN00187

A/Experimental source: strain C58 (Dupont)

C/Genetics:

A/Map position: linear chromosome

Query Match 12.0%; Score 93.5; DB 2; Length 145;

Best Local Similarity 21.8%; Pred. No. 0.48;

Matches 34; Conservative 25; Mismatches 64; Indels 33; Gaps 4;

QY 3 LKVAAPAAIVVSGSLAGVVPQMG-----GNNHGGNSGPDSTLTIYQGSANA 55

Db 1 MIRKSFIALVALVGLSAAAPAMANDVRIEYQGSNSAGAGQEGGNRIKRTYQNGYNR 60

QY 56 ALALQSDARKSETTITQSGYNGADYQ--GADNYDQLVTRVVTHEMAHADQNAKNSDIT 114

Db 61 IVG-----HGYRHNLISAVGQEGHDNYGSTTQ-----NGNRNVAG 95

QY 115 VQYGGNNALVNOTASDSSVWVROYGFGNNATANO 150

Db 96 IGFQFSNHTTILTDGNGNIAAGVQYGRGCSANVSQ 131

RESULT 13

H98144

hypothetical protein AGR_L_228 [imported] - Agrobacterium tumefaciens (strain C58, Cerec

C/Species: Agrobacterium tumefaciens

C/Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002

C/Accession: H98144

R./Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,

A.; Liu, F.; Woliam, C.; Allinger, M.; Doughy, D.; Scott, C.; Lappae, C.; Markelz, B.;

Science 294, 2323-2328, 2001

A/Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum

A/Reference number: A97359; PMID:21608551; PMID:11743194

A/Accession: H98144

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-145 <RUR>

A/Cross-references: GB:AE007870; PIDN:AAK8682.1; PID:g15158413; GSPDB:GN00170

C/Genetics:

A/Map position: linear chromosome

Query Match 12.0%; Score 93.5; DB 2; Length 145;

Best Local Similarity 21.8%; Pred. No. 0.48;

Matches 34; Conservative 25; Mismatches 64; Indels 33; Gaps 4;

QY 3 LKVAAPAAIVVSGSLAGVVPQMG-----GNNHGGNSGPDSTLTIYQGSANA 55

Db 1 MIRKSFIALVALVGLSAAAPAMANDVRIEYQGSNSAGAGQEGGNRIKRTYQNGYNR 60

QY 56 ALALQSDARKSETTITQSGYNGADYQ--GADNYDQLVTRVVTHEMAHADQNAKNSDIT 114

Db 61 IVG-----HGYRHNLISAVGQEGHDNYGSTTQ-----NGNRNVAG 95

QY 115 VQYGGNNALVNOTASDSSVWVROYGFGNNATANO 150

Db 96 IGFQFSNHTTILTDGNGNIAAGVQYGRGCSANVSQ 131

RESULT 14

F70825

probable PPE protein - Mycobacterium tuberculosis (strain H37Rv)

C/Species: Mycobacterium tuberculosis

C/Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 15-Sep-2003

C/Accession: F70825

R./Cole, S.T.; Broesch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon,

A.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A/Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A/Reference number: A70500; PMID:98295987; PMID:9634230

A/Accession: F70825

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-645 <COL>

A/Cross-references: GB:AL021958; GB:AL123456; NID:g3261536; PIDN:CAA17522.1; PID:e12532

A/Experimental source: strain H37Rv

C/Genetics:

A/Map position: PPE

Query Match 12.0%; Score 93.5; DB 2; Length 645;

Best Local Similarity 24.5%; Pred. No. 2.6;

Matches 39; Conservative 17; Mismatches 68; Indels 35; Gaps 6;

QY 15 SCSALAGVVPQ-----WGGGNNHGGNSGPDSTLTIYQGSANALALQSDARKSETT 69

Db 237 SCNGNVGICPSSFPYQSGNIGANVGGNSG--DNNFGFNPFGNANIGNAGPNNSSPAV 295

QY 70 ITQSG---GYNGADYQGADNYDQLVTRVVTHEMAHADQNAKNSDITVQYG----- 119

Db 296 PTPGNGNVGICGNGGNGNFGCGNTGN-----ANIGLVGVGDGVGFGNSGYNFGP 345

QY 120 ---GNNAALVNOTASDS-----SVWVROYGFGNNATANO 149

Db 346 GNTGNNNIGIGILGTSNQGIFGGLNGSGNIGFGNSGTGN 384

RESULT 15

S11672

ice nucleation protein - Xanthomonas campestris

C/Species: Xanthomonas campestris

C/Date: 21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 26-Aug-1999

C/Accession: S11672

R./Zhao, J.; Orser, C.S.

Mol. Gen. Genet. 223, 163-166, 1990

A/Title: Conserved repetition in the ice nucleation gene inax from Xanthomonas campestris

A/Reference number: S11672; PMID:91080859; PMID:2259339

A/Accession: S11672

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-1567 <ZHA>

A/Cross-references: EMBL:X52970; NID:g48531; PIDN:CAA37140.1; PID:g48532

C/Suprafamily: ice nucleation protein

Query Match 12.0%; Score 93.5; DB 2; Length 1567;

Best Local Similarity 28.2%; Pred. No. 6.9;

Matches 42; Conservative 21; Mismatches 41; Indels 45; Gaps 10;

QY 34 GGGNSGPDSTLTIYQGSANALALQSDARKSETTITQSGYNGADYQGAD---NY-- 88

```

Db      383  GSTGTAGADSTL-IAGYSTQTA-----GSESLT-AGYSTQTARQSDITAGYGS 432
Oy      89   -----DQVTRVVTHEMAHD-----OMNAKSDITVQYG-----GNNALV 126
Db      433  TGTAGADSTL IAGYSTQTSQSDSLT IAGYSTQTARKSDITAG-YGSTGTAGSDSLI 491
Oy      127  -----NOTASDSSVVRQVGFNNATANQ 150
Db      492  AGYSTQTAGSESLT--AGYSTQTAQ 518

```

Search completed: March 11, 2004, 18:42:10
 Job time : 10.3 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:14:48 ; Search time 6.3 Seconds

(without alignments)
1248.031 Million cell updates/sec

Title: US-09-543-407-22

Perfect score: 776
Sequence: 1 MKLLKVAFAAIVSGSALA.....DSSVMRVQVFGNNATANY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	682	87.9	151	1	CSGA_SALTY
2	521	67.1	151	1	CSGA_SALTY
3	499.5	64.4	152	1	CSGA_SALTY
4	93.5	12.0	1567	1	ICEN_XANCT
5	92.5	11.9	151	1	CSGB_ECOLI
6	91.5	11.8	1034	1	ICEN_PANAN
7	91.5	11.8	1258	1	ICEN_ERHME
8	91	11.7	1754	1	PMBB_CHUTR
9	90.5	11.7	151	1	CSGB_SALTY
10	90.5	11.7	151	1	CSGB_SALTY
11	89.5	11.5	1332	1	ICBA_PANAN
12	87	11.2	491	1	YK98_MYCTU
13	87	11.2	580	1	GP63_LEIDO
14	87	11.2	539	1	GP63_LEICH
15	87	11.2	602	1	GP63_LEIMA
16	87	11.2	646	1	VP38_LEIME
17	86.5	11.1	262	1	VG38_BPT2
18	86.5	11.1	487	1	Y442_MYCTU
19	86.5	11.1	548	1	CEAK_ECOLI
20	86.5	11.1	1028	1	OVO_DROME
21	85.5	11.0	1210	1	ICEN_PSEFL
22	85	11.0	1778	1	N189_SCHPO
23	84.5	10.9	495	1	PO33_MOUSE
24	84.5	10.9	760	1	YBIL_ECOLI
25	84.5	10.9	1211	1	BUN2_DROME
26	83.5	10.8	497	1	PO33_PAT
27	83	10.7	1140	1	YV96_YEAST
28	82	10.6	500	1	PO33_HUMAN
29	82	10.6	392	1	YF48_MYCTU
30	81.5	10.5	698	1	HME1_HUMAN
31	81.5	10.5	737	1	YD76_MYCPN
32	81.5	10.5	2038	1	FSH_DROME
33	81	10.4	165	1	GRP1_ORYSA

ALIGNMENTS

RESULT 1	CSGA_SALTY	STANDARD	PRT	151 AA.	
AC	P55225;				Q9ZFHO azotobacter
DT	01-OCT-1996 (Rel. 34, Created)				O33479 pseudomonas
DT	01-OCT-1996 (Rel. 34, Last sequence update)				O07202 medicago sa
DT	10-OCT-2003 (Rel. 42, Last annotation update)				O05156 staphylococ
DE	Major curlin subunit precursor (Fimbrin SEF17).				P18899 saccharomyc
GN	CSGA OR AGFA OR STM144 OR STY1181 OR T1776.				O10540 mycobacteri
OS	Salmonella typhimurium,				Q50630 mycobacteri
OS	Salmonella typhi, and				P91698 drosophila
OS	Salmonella enteritidis.				P51989 xenopus lae
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;				O30611 pseudomonas
OC	Enterobacteriaceae; Salmonella.				P49687 saccharomyc
OX	NCBI_TaxID=602, 601, 592;				Q50367 mycoplasma
RP	SEQUENCE FROM N.A.				
RC	SPECIES=S. typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;				
RC	MEDLINE=98117058; PubMed=9457880;				
RA	Romling U., Bian Z., Hammar M., Sierralta W.D., Normark S.;				
RT	"Curli fibers are highly conserved between Salmonella typhimurium and				
RT	Escherichia coli with respect to operon structure and regulation.";				
RL	J. Bacteriol. 180:722-731(1998).				
RP	SEQUENCE FROM N.A.				
RC	SPECIES=S. typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;				
RC	MEDLINE=21534948; PubMed=11677609;				
RA	McClintock M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,				
RA	Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,				
RA	Leonard S., Nguyen C., Scott K., Holmes A., Grevell N., Mulvaney E.,				
RA	Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,				
RA	Waterston R., Wilson R.K.;				
RT	"Complete genome sequence of Salmonella enterica serovar Typhimurium				
RT	LT2.";				
RL	Nature 413:852-856(2001).				
RP	SEQUENCE FROM N.A.				
RC	SPECIES=S. typhi; STRAIN=CT18;				
RC	MEDLINE=21534947; PubMed=11677608;				
RA	Parthill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,				
RA	Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebatia M.,				
RA	Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,				
RA	Cronin A., Davies P., Davies R.M., Dowd L., White N., Farrer J.,				
RA	Fellwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,				
RA	Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,				
RA	Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,				
RA	Whitehead S., Barrett B.G.;				
RT	"Complete genome sequence of a multiple drug resistant Salmonella				
RT	enterica serovar Typhi CT18.";				
RL	Nature 413:848-852(2001).				
RP	SEQUENCE FROM N.A.				
RC	SPECIES=S. typhi; STRAIN=Ty2 / ATCC 700931;				
RC	MEDLINE=22531367; PubMed=12644504;				
RA	Deng W., Lion S.R., Plunkett G. III, Mayhew G.F., Rose D.J.,				
RA	Burford V., Kodoyanni V., Schwartz D.C., Blattner F.R.;				

CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: L04979; AAA3616.1; -
 DR EMBL: X80754; CAA5282.1; -
 DR EMBL: AE000205; AAC74126.1; -
 DR EMBL: D90741; BAA35832.1; -
 DR EMBL: D90742; BAA35840.1; -
 DR PIR: S70788; S70788.
 DR EcoGene: EG11489; CSGA.
 KW Fimbrin; Signal; Complete proteome.
 FT SIGNAL 1 20
 FT CHAIN 1 151 MAJOR CURLIN SUBUNIT.
 FT CONFLICT 7 7 A -> B (IN REF. 1).
 FT CONFLICT 7 7 A -> B (IN REF. 1).
 SQ SEQUENCE 151 AA; 15049 MW; C003470D208D395F CRC64;
 Query Match 67.1%; Score 521; DB 1; Length 151;
 Best Local Similarity 69.5%; Pred. No. 6e-37;
 Matches 105; Conservative 17; Mismatches 29; Indels 0; Gaps 0;
 QY 1 MLKLKVAFAFAIVSGSALAGVVPWVGSGGNGHGGSGPDSLTISIYOGSANAALALQ 60
 DB 1 MLKLKVAFAFAIVSGSALAGVVPWVGSGGNGHGGSGNSPSELNITIOYGGNSALALQ 60
 QY 61 SPARKSETTITQSGYGNGADVGGADNYDQLVTRVVTHEMAHADQMAKNSDITVQGYG 120
 DB 61 TDARNSDLITIQHGGNGADVGGSDSDSLTLQRFNGSATLDDQNGKSEMTVQVFGG 120
 QY 121 NNALVNOTASDSSVMVRFVGFNNATANOY 151
 DB 121 GNGAADVDTASNSVTVTVGFNNATANOY 151
 RESULT 3
 CSGA_ECO57 STANDARD; PRT; 152 AA.
 AC 093024;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Major curlin subunit precursor.
 OS CSGA OR Z1676 OR ECS1420.
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=83334;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / ATCC 43895;
 RX MEDLINE=21218556; PubMed=11319125;
 RA Uhlich G.A., Keen J.E., Elder R.O.;
 RT "Mutations in the csdD promoter associated with variations in curli
 RT expression in certain strains of Escherichia coli O157:H7.";
 RL Appl. Environ. Microbiol. 67:2367-2370(2001).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Ferna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
 RA Apodoca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
 RL Nature 409:529-533(2001).

RN (3)
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RIMD 0509952;
 RX MEDLINE=21156211; PubMed=11258796;
 RA Hayashi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22(2001).
 CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 CC FIBRONECTIN.
 CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF275733; AAK53212.1; -
 DR EMBL: AE005315; AAG55788.1; -
 DR EMBL: AP002554; BAB34843.1; -
 DR PIR: D90806; D90806.
 DR PIR: H85665; H85665.
 KW Fimbrin; Signal; Complete proteome.
 FT SIGNAL 1 20
 FT CHAIN 1 152 MAJOR CURLIN SUBUNIT.
 SQ SEQUENCE 152 AA; 15099 MW; EE2D2D94DDE91243 CRC64;
 Query Match 64.4%; Score 499.5; DB 1; Length 152;
 Best Local Similarity 67.8%; Pred. No. 3.8e-35;
 Matches 103; Conservative 18; Mismatches 30; Indels 1; Gaps 1;
 QY 1 MLKLKVAFAFAIVSGSALAGVVPWVGSGGNGHGGSGPDSLTISIYOGSANAALAL 59
 DB 1 MLKLKVAFAFAIVSGSALAGVVPWVGSGGNGHGGSGNSPSELNITIOYGGNSALAL 60
 QY 60 OSDARKSETTITQSGYGNGADVGGADNYDQLVTRVVTHEMAHADQMAKNSDITVQGYG 119
 DB 61 QDARNSDLITIQHGGNGADVGGSDSDSLTLQRFNGSATLDDQNGKSDIMTVQKRG 120
 QY 120 GNNALVNOTASDSSVMVRFVGFNNATANOY 151
 DB 121 GNGAADVDTASNSTVTVTVGFNNATANOY 152
 RESULT 4
 ICEN_XANCT STANDARD; PRT; 1567 AA.
 AC P18127;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Ice nucleation protein.
 GN INAX.
 OS Xanthomonas campestris (pv. translucens).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 OC Xanthomonadaceae; Xanthomonas.
 OX NCBI_TaxID=343;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=X565;
 RX MEDLINE=91080659; PubMed=2259339;
 RA Zhao J., Orser C.S.;
 RT "Conserved repetition in the ice nucleation gene inax from
 RT Xanthomonas campestris pv. translucens.";
 RL Mol. Gen. Genet. 223:163-166(1990).

CC - FUNCTION: Ice nucleation proteins enable bacteria to nucleate
 CC crystallization in supercooled water.
 CC SUBCELLULAR LOCATION: Outer membrane (By similarity).
 CC - DOMAIN: CONTAINS 153 IMPERFECT REPEATS OF THE CONSENSUS
 CC OCTAPEPTIDE A-G-Y-G-S-T-L-T. FURTHER ON A 16-RESIDUE AND A
 CC REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.
 CC MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE
 CC NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.
 CC - SIMILARITY: Belongs to the bacterial ice nucleation protein
 CC family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X52970; CAA37140.1; -.
 DR HSSP: P06620; 11NA.
 DR InterPro: IPR000258; Ice nucleatn.
 DR Pfam: PF00818; Ice nucleation; 81.
 DR PRINTS: PR00327; ICENUCLEATN.
 DR PROSITE: PS00314; ICE_NUCLEATION; 57.
 DR Ice nucleation; Repeat; Outer membrane.
 KW SEQUENCE 1567 AA; 152548 MW; C8B451D959ECAD63 CRC64;
 SQ
 Query Match 12.0%; Score 93.5; DB 1; Length 1567;
 Best Local Similarity 28.2%; Pred. No. 4;
 Matches 42; Conservative 21; Mismatches 41; Indels 45; Gaps 10;
 QY 34 GGAGSSGPDSTLSTIYQGSANALALQSDARKSETTITQSGYGNAGVQCAD---NY-- 88
 DB 363 GSYTAGADSTL- IAGYSTQTA-----GGSSSLT-AGYGTQARQSDITAYGS 432
 QY 89 -----DQLVRYVTHMAHAD-----QMNKNSITTYGQY---GNNALV 126
 DB 433 TGTAGADSTL IAGYSTQTSQSDSLNAGYSTYARQSDITAG-YGSTGTAGSDSLI 491
 QY 127 -----NOTASDSVVRQVGFGNATANO 150
 DB 492 AGYGTQARQSDSLT-AGYGTQTAQ 518

RESULT 5
 CSGB_ECOLI STANDARD; PRT; 151 AA.
 ID CSGB_ECOLI STANDARD; PRT; 151 AA.
 AC P39828;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Minor curliin subunit precursor.
 GN CSGB OR B1041 OR Z1675 OR ECS1419.
 OS Escherichia coli, and
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562, 83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MC4100;
 RX MEDLINE=9641468; PubMed=8817489;
 RA Hammar M., Arngvist A., Bian Z., Olsen A., Normark S.;
 RT "Expression of two *csa* operons is required for production of
 RT fibrinectin- and congo red-binding curli polymers in *Escherichia coli*
 RT K-12.";
 RT Mol. Microbiol. 18:661-670 (1995).
 RL [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=9742617; PubMed=9278503;
 RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,

RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of *Escherichia coli* K-12.";
 RL Science 277:1453-1474 (1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97061202; PubMed=8905232;
 RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
 RA Mori H., Moromura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
 RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horiiuchi T.;
 RT "A 7.8-Kb DNA sequence of the *Escherichia coli* K-12 genome
 RT corresponding to the 12.7-28.0 min region on the linkage map.";
 RL DNA Res. 3:137-155 (1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blatner F.R.;
 RT "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7.";
 RL Nature 409:529-533 (2001).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasekawa C., Ogasawara N., Yasunaga T.,
 RA Kubura S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohaemorrhagic *Escherichia coli*
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:111-22 (2001).
 RN [6]
 RP SEQUENCE OF 1-21 FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=95157246; PubMed=7854117;
 RA Arngvist A., Olsen A., Normark S.;
 RT "Sigma S-dependent growth-phase induction of the *csaBA* promoter in
 RT *Escherichia coli* can be achieved in vivo by sigma 70 in the absence
 RT of the nucleoid-associated protein H-NS.";
 RL Mol. Microbiol. 13:1021-1032 (1994).
 CC - FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
 CC CURLIN MONOMERS.
 CC -----
 CC - SIMILARITY: BELONGS TO THE CSGB/CSGB FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X90754; CAA62281.1; -.
 DR EMBL: AE000205; AAC74125.1; -.
 DR EMBL: D90741; BAA35831.1; -.
 DR EMBL: AE005315; AAG55787.1; -.
 DR EMBL: AF002554; BAB34842.1; -.
 DR PIR: C90806; G90806.
 DR PIR: G85665; G85665.
 DR PIR: S70787; S70787.

DR EcGene; EGI2621; cagB.
 KW Fimbrin; Signal; Complete proteome.
 FT SIGNAL 1 21 POTENTIAL.
 CHAIN 22 151 MINOR CURLIN SUBUNIT.
 SQ SEQUENCE 151 AA; 15882 MW; B18D26B964014B8 CRC64;

Query Match 11.8%; Score 92.5; DB 1; Length 151;
 Best Local Similarity 28.1%; Pred. No. 0.36;
 Matches 32; Conservative 15; Mismatches 62; Indels 5; Gaps 2;

QY 38 SSGPSTLSIYOGSANAALALQSDARKSETTITGSGYGAGDVGADNDVDTLYRVYT 97
 DB 21 AAGYDLANSRYN-----AVNELSKSFNQAATIGQAGTNSAQLRQGGSLAIVAAQEGS 76

QY 98 HEMAHADQWNAKNSDITVGOYGGNNAALVNOTASDSSVWVROYGFGNNAATANQY 151
 DB 77 SNRAKIDQGDVYL-AVIDQAGSANDASISGAYGVTAMTIQSGSNKANITGY 129

RESULT 6
 ICEN PANAN STANDARD; PRT; 1034 AA.
 AC 047879;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Ice nucleation protein Inau.
 GN INAU.
 OS Pantoea ananas (Erwinia uredovora).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Pantoea.
 OX NCBI_TaxId=553;
 RP [1]
 RN SEQUENCE FROM N.A.
 RX STRAIN-KUIN-3:
 RC MEDLINE=94264407; PubMed=7764866;
 RA Michigant Y., Watabe S., Abe K., Obara H., Arai S.;
 RT "Cloning and sequencing of an ice nucleation active gene of Erwinia
 RT uredovora".
 RL Bioest. Biochem. 58:762-764(1994).
 CC -1- FUNCTION: Ice nucleation proteins enable bacteria to nucleate
 CC crystallization in supercooled water.
 CC -1- SUBCELLULAR LOCATION: Outer membrane.
 CC -1- DOMAIN: CONTAINS IMPERFECT REPEATS OF A CONSENSUS OCTAPEPTIDE
 CC A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A REGIONAL 48-RESIDUE
 CC PERIODICITY IS SUPERIMPOSED.
 CC -1- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE
 CC NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.
 CC -1- SIMILARITY: Belongs to the bacterial ice nucleation protein
 CC family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; D14992; BAA03636.1; -.
 DR PIR; JC2143; JC2143.
 DR HSSP; P06620; 11NA.
 DR InterPro; IPR000258; Ice nucleatn.
 DR Pfam; PF00818; Ice nucleatn; 51.
 DR PRINTS; PR00327; ICENUCLEATN
 DR PROSITE; PS00314; ICE_NUCLEATION; 34.
 KW Ice nucleation; Repeat; Outer membrane.
 FT DOMAIN 162 993 OCTAPEPTIDE PERIODICITY.
 SQ SEQUENCE 1034 AA; 103378 MW; FA222523D333EAD CRC64;

Query Match 11.8%; Score 91.5; DB 1; Length 1034;
 Best Local Similarity 26.3%; Pred. No. 3.7;
 Matches 46; Conservative 20; Mismatches 54; Indels 55; Gaps 10;

QY 12 IVVSGSALAGVVPQW--GGGNHNGGNS-----GPDSTLSIYOGSANAAL 57
 DB 161 IATYGSSTLSGTHQSOLIAIGYSTETPAGDSSTLIAGYSTGTAGSDSTL-VAGGSGTQTA- 218

QY 58 ALQSDARKSETTITGSGYGN-----GADVGOG-----ADVNDOLVTRVTHEMAHAD- 104
 DB 219 -----GEESQWAGYSGTQTKMGKSDLTGAGYSTGTAGDSSLIAGYSGTQTAGEDS 270

QY 105 -----QWNAKNSDITVGOYGGNNAALVNOTASDSSVWVROYGFGNNAATANQ 150
 DB 271 SLTAGYSTQTKQKSDLTAG--YGSTGTA-----GADSSLI--AGYSGTQTAG 316

RESULT 7
 ICEN ERWHE STANDARD; PRT; 1258 AA.
 AC P16239;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Ice nucleation protein.
 GN ICEB.
 OS Erwinia herbicola.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Pantoea.
 OX NCBI_TaxId=549;
 RP [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=M1;
 RX MEDLINE=90152370; PubMed=2515997;
 RA Warren G.J., Corotto L.V.;
 RT "The consensus sequence of ice nucleation proteins from Erwinia
 RT herbicola, Pseudomonas fluorescens and Pseudomonas syringae".
 RL Gene 85:239-242(1989).
 CC -1- FUNCTION: Ice nucleation proteins enable bacteria to nucleate
 CC crystallization in supercooled water.
 CC -1- SUBCELLULAR LOCATION: Outer membrane.
 CC -1- DOMAIN: CONTAINS 126 IMPERFECT REPEATS OF A CONSENSUS OCTAPEPTIDE
 CC A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A REGIONAL 48-RESIDUE
 CC PERIODICITY IS SUPERIMPOSED.
 CC -1- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE
 CC NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.
 CC -1- SIMILARITY: Belongs to the bacterial ice nucleation protein
 CC family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; M26382; AAA24823.1; -.
 DR PIR; JQ0188; JQ0188.
 DR HSSP; P06620; 11NA.
 DR InterPro; IPR000258; Ice nucleatn.
 DR Pfam; PF00818; Ice nucleatn; 65.
 DR PRINTS; PR00327; ICENUCLEATN
 DR PROSITE; PS00314; ICE_NUCLEATION; 45.
 KW Ice nucleation; Repeat; Outer membrane.
 FT DOMAIN 162 1217 OCTAPEPTIDE PERIODICITY.
 SQ SEQUENCE 1258 AA; 125084 MW; 590EBA130077FBD4 CRC64;

Query Match 11.8%; Score 91.5; DB 1; Length 1258;
 Best Local Similarity 26.3%; Pred. No. 4.6;
 Matches 46; Conservative 20; Mismatches 54; Indels 55; Gaps 10;

QY 12 IVVSGSALAGVVPQW--GGGNHNGGNS-----GPDSTLSIYOGSANAAL 57
 DB 161 IATYGSSTLSGTHQSOLIAIGYSTETPAGDSSTLIAGYSTGTAGABDTL-VAGGSGTQTA- 218

Qy	58	ALQSDARKSTLTITQSSVGN-----GAIVVGG-----ADVYDPLVIRVVYTHMAAD- 104
Db	219	-----GESSQMGVGYSTQTKMGKSDLTLAGYSTGTAQDSSLINGVYSTQTAAEDS 270
Qy	105	-----QMNKNSDITVQGYGCKNNALVWQTASDSSVMYKQVGFENNAATNQ 150
Db	271	SLTAGVYSTQTAAQKSDLING-VGSGTGA-----GADSSLI---AGVYSTQTAAE 316

RESULT & DISCUSSION

ID	PMDB CHLTR	STANDARD;	PRT;	1754 AA.
AD	084418;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Probable outer membrane protein pmpB precursor (Polymorphic membrane protein B).			
DE	PMPB OR CT413.			
GN	Chlamydia trachomatis.			
OS	Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.			
OX	NCBI_Taxid=813;			
RN	(1)			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=D/UW-3/CX.			
RX	MEDLINE=99000809; PubMed=9784136;			
RA	Stephens R.S., Kalmán S., Lammel C.J., Fan J., Marathe R., Aravind L., Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V., Davis R.W.,			
RT	"Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis."			
RL	Science 282:754-759(1998).			
CC	-1 SUBCELLULAR LOCATION: Cell wall surface (elementary bodies) (Potential).			
CC	-1 SIMILARITY: Belongs to the pmp outer membrane protein family.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way CC modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/)			
CC	or send an email to license@sib.ch).			
CC	-----			
DR	EMBL, AB001314; AAC68010.1; AIT_INIT.			
DR	PfCI-2DPAGE; 084418; -.			
DR	InterPro; IPR003368; Chlamydia_PMP.			
DR	Pfam; PF02415; Chlamydia_PMP; 4.			
DR	TIGRFAMs; TIGR01376; POMP repeat; 18.			
KW	Outer membrane; Signal; Multigene family; Complete proteome.			
FT	SIGNAL	1	1754	POTENTIAL.
FT	CHAIN	1	1754	POTENTIAL.
Q	SEQUENCE	1754 AA;	183317 MW;	9696CF8085D36185D CRC64;

ID	CSGB_SALT1	STANDARD;	PRT;	151 AA.
AC	0827M3;			
DT	28-FEB-2003 (Rel.. 41, Created)			
DT	28-FEB-2003 (Rel.. 41, Last sequence update)			
DT	10-OCT-2003 (Rel.. 42, Last annotation update)			
DE	Minor curlin subunit precursor.			
GN	CSGB OR STY1180 OR T1177.			
OS	Salmonella typhi.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OC	Enterobacteriaceae; Salmonella.			
OX	NCBI_Taxid=601;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CT18;			
RX	MEDLINE=21534947; PubMed=11677608;			
RA	Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,			
RA	Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,			
RA	Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,			
RA	Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,			
RA	Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagers K.,			
RA	Krogh A., Larsen T.S., Leather S., Moulé S., O'Goira P., Parry C.,			
RA	Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,			
RT	Whitehead S., Barrrell B.G.;			
RT	"Complete genome sequence of a multiple drug resistant <i>Salmonella</i>			
RU	<i>enterica</i> serovar Typhi CT18.";			
RU	Nature 413:848-852(2001).			

РЕСПУБЛИКА

CSGB_SALTI

RESULT 10

CSGB_SALTY	STANDARD;	PRT;	151 AA
------------	-----------	------	--------

RESULT 12
 YK98 MYCTU STANDARD; PRT; 491 AA.
 ID YK98 MYCTU STANDARD; PRT; 491 AA.
 AC 01-OCT-1996 (Rel. 34, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Hypothetical glycine-rich protein Rv2098c/MT2159/MD2125c.
 GN Rv2098c OR MT2159 OR MTCY49.38c OR MB2125c.
 OS Mycobacterium tuberculosis, and
 OC Mycobacterium bovis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
 OC NCBI_TaxID=1773, 1765;
 [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=M.tuberculosis; STRAIN=H37Rv;
 RX MEDLINE=9829587; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
 Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
 Hornsby T., Jorgels K., Krogh A., McLean J., Moule S., Murphy J.,
 Oliver S., Osborne U., Quail M.A., Rajandream M.A., Rogers R.,
 Sulston J.E., Taylor K., Whitehead S., Squares S., Rogers R.,
 RA Rulster S., Seeger K., Skelton S., Squares S., Rogers R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.,
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 RL Nature 393:537-544 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
 RX MEDLINE=22206494; PubMed=12218036;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 Peterson J., Deboy R., Dodson R., Gwinn M., Hatt D., Hickey R.,
 Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
 Balch A., Utterback T., Weidman J., Khouli H., Gill U., Mikula A.,
 Bishel W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.,
 RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains."
 RL J. Bacteriol. 184:5479-5490 (2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC SPECIES=M.bovis; STRAIN=AF2122/97;
 RX MEDLINE=22709107; PubMed=12788972;
 RA Garnier T., Eigmeier K., Camus J.-C., Medina N., Mansoor H.,
 Pryor M., Duthoy S., Grondin S., Jacroix C., Monsemp C., Simon S.,
 Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
 Parkhill J., Barrett B.G., Cole S.T., Gordon S.V., Hewinson R.G.,
 RT "The complete genome sequence of Mycobacterium bovis."
 RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882 (2003).
 CC -1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGMS
 SUBFAMILY.
 CC -1- CAUTION: Ref.1 sequence differs from that shown due to a
 CC frameshift in position 59. Ref.1 sequence has been checked by
 CC authors in Ref.1 and they report that no errors have been found.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: Z73966; CA98228.1; ALT_FRAME.
 DR EMBL: AEO07065; AAK6440.1; -
 DR EMBL: BX248341; CAD96978.1; -
 DR TIGR: MT2159; -
 DR InterPro: IPR000084; PE_region.
 DR Pfam: PF00934; PE; 1.

DR ProDom: PD001223; PE_region: 1.
 KW Hypothetical protein; Complete proteome.
 FT CONFLICT 312 312 G -> GG (IN REF. 1).
 SQ SEQUENCE 491 AA; 41979 MW; 12C8630C59CA0C13 CRC64;
 Query Match 11.2%; Score 87; DB 1; Length 491;
 Best local Similarity 28.2%; Pred. No. 3.8; Indels 26; Gaps 4;
 Matches 31; Conservative 8; Mismatches 45;
 QY 20 AGVPMQGGGN-----HNGGNSGPDSTLSIYQYSANMALAQSDARKSETTITQSG 74
 DB 222 AGLIGHGAGAGNGDGHGSGSGAGSGSGSGGFGGAGGL----- 264
 QY 75 YGNGAVGAGADYDLYRVYTHMAHMDQNMANKSD---ITVQYGN 121
 DB 265 YGNGAAGSGAGNGD-AGTGVSSDGFALGSGSGRGDGLIGVGGG 313
 RESULT 13
 GP63 LEIDO STANDARD; PRT; 590 AA.
 ID GP63 LEIDO STANDARD; PRT; 590 AA.
 AC P23223;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Leishmanolysin precursor (EC 3.4.24.36) (Cell surface protease)
 DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface
 DE endopeptidase).
 GN GP63.
 OS Leishmania donovani.
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 OC NCBI_TaxID=5661;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LV9;
 RX MEDLINE=92107220; PubMed=1762629;
 RA Webb J.R., Buton L.L., McMaster R.W.,
 RT "Heterogeneity of the genes encoding the major surface glycoprotein
 RT of Leishmania donovani".
 RL Mol. Biochem. Parasitol. 48:173-184 (1991).
 CC -1- FUNCTION: Has an integral role during the infection of macrophages
 CC in the mammalian host.
 CC -1- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and
 CC P1', and basic residues at P2 and P3'. A model nonapeptide is
 CC cleaved at -Ala-Tyr-|-Leu-Lys-Lys-|-.
 CC -1- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
 CC -1- SIMILARITY: Belongs to peptidase family M8.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M60048; AAA29244.1; -
 DR HSSP: P08148; LMLL.
 DR MEROPS: M08.001; -
 DR InterPro: IPR006025; Pept_M_Zn_BS.
 DR InterPro: IPR001577; Peptidase_M8.
 DR Pfam: PF01457; Peptidase_M8; 1.
 DR PRINTS: PR00782; LSHMANOLYSIN.
 DR PROSITE: PS00142; ZINC_PROTEASE; 1.
 KW Hydroxylase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;
 KM Zymogen; Signal; Cell adhesion; GPI-anchor; Lipoprotein.
 FT SIGNAL 1 39
 FT PROPEP 87
 FT CHAIN 40 565
 FT PROPEP 566 590
 FT METAL 251 251
 FT ACT_SITE 252 252
 FT BY SIMILARITY.

FT METAL 255 255 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 321 321 ZINC (CATALYTIC) (BY SIMILARITY).
 FT DISULFID 112 129 BY SIMILARITY.
 FT DISULFID 178 217 BY SIMILARITY.
 FT DISULFID 301 373 BY SIMILARITY.
 FT DISULFID 380 443 BY SIMILARITY.
 FT DISULFID 393 412 BY SIMILARITY.
 FT DISULFID 402 477 BY SIMILARITY.
 FT DISULFID 454 498 BY SIMILARITY.
 FT DISULFID 503 553 BY SIMILARITY.
 FT DISULFID 523 546 BY SIMILARITY.
 FT CARBOHYD 287 287 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT LIPID 565 565 GPI-anchor amidated asparagine (By similarity).
 SQ SEQUENCE 590 AA; 62950 MW; 0FB315D29659F58 CRC64;
 Query Match 11.2%; Score 87; DB 1; Length 590;
 Best Local Similarity 89.5%; Pred. No. 4.7;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 85 ADNYDQLVTRVVTHEMAHA 103
 DB 238 ASRYDQLVTRVVTHEMAHA 256
 RESULT 14
 GP63 LEICH STANDARD; PRT; 599 AA.
 ID GP63 LEICH
 AC P15706;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Leishmanolysin precursor (EC 3.4.24.36) (Cell surface protease)
 DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface endopeptidase).
 GN GP63.
 OS Leishmania chagasi.
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 OX NCBI_TaxId=44271;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90205976; PubMed=2320059;
 RA Miller R.A., Reed S.G., Parsons M.;
 RT "Leishmania gp63 molecule implicated in cellular adhesion lacks an Arg-Gly-Asp sequence."
 RL Mol. Biochem. Parasitol. 39:267-274(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92112918; PubMed=1370484;
 RA Ramamoorthy R., Donelson J.E., Paetz K.E., Maybodi M., Roberts S.C., Wilson M.E.;
 RT "Three distinct RNAs for the surface protease gp63 are differentially expressed during development of Leishmania donovani chagasi promastigotes to an infectious form."
 RL J. Biol. Chem. 267:1888-1895(1992).
 CC -1- FUNCTION: Has an integral role during the infection of macrophages in the mammalian host.
 CC -1- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and P1' and basic residues at P2 and P3'. A model nonapeptide is cleaved at -Ala-Tyr-[Leu-Lys-Lys-].
 CC -1- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
 CC -1- SIMILARITY: Belongs to peptidase family M8.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation-CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC EMBL; M80672; AAA29238.1; -.

DR EMBL; M28527; AAA29235.1; -.
 DR PIR; A44951; A44951.
 DR HSSP; P08148; 1LML.
 DR MEROPS; M08.001; -.
 DR InterPro; IPR006025; Pept M Zn BS.
 DR InterPro; IPR001577; Peptidase_M8.
 DR Pfam; PF01457; Peptidase_M8; 1.
 DR PRINTS; PR00782; LSHMANOLYSIN.
 DR PROSITE; PS00142; ZINC PROTEASE; 1.
 DR HydroLase; Metalloprotease; Glycoprotein; Metal-binding; Zinc; KX
 KW Hydrogen, Signal; Cell adhesion; GPI-anchor; Lipoprotein.
 FT SIGNAL 1 39
 FT PROPEP 40 97
 FT CHAIN 98 574
 FT PROPEP 575 599
 FT METAL 261 261
 FT ACT SITE 262 262
 FT METAL 265 265
 FT METAL 331 331
 FT DISULFID 122 139
 FT DISULFID 188 227
 FT DISULFID 311 383
 FT DISULFID 390 452
 FT DISULFID 403 422
 FT DISULFID 412 486
 FT DISULFID 463 507
 FT DISULFID 512 562
 FT DISULFID 532 555
 FT CARBOHYD 297 394
 FT CARBOHYD 394 394
 FT LIPID 574 574
 SQ SEQUENCE 599 AA; 63848 MW; 746730AE8E2A2E7C CRC64;
 Query Match 11.2%; Score 87; DB 1; Length 599;
 Best Local Similarity 89.5%; Pred. No. 4.8;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 85 ADNYDQLVTRVVTHEMAHA 103
 DB 248 ASRYDQLVTRVVTHEMAHA 266
 RESULT 15
 GP63 LEIMA STANDARD; PRT; 602 AA.
 ID GP63 LEIMA
 AC P08148; P15906;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Leishmanolysin precursor (EC 3.4.24.36) (Cell surface protease)
 DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface endopeptidase).
 GN GP63.
 OS Leishmania major.
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 OX NCBI_TaxId=5664;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 101-123.
 RX MEDLINE=88154764; PubMed=3346625;
 RA Buton L.L., McMaster W.R.;
 RT "Molecular cloning of the major surface antigen of Leishmania."
 RL J. Exp. Med. 167:724-729(1988).
 RN [2]
 RP REVISIONS.
 RA Buton L.L., McMaster W.R.;
 RL J. Exp. Med. 171:589-589(1990).
 RN [3]
 RP GPI-ANCHOR.
 RX MEDLINE=91009116; PubMed=2145267;
 RA Schneider P., Ferguson M.A.J., McConville M.J., Mehler A., Homans S.W., Border C.;
 RT "Structure of the glycosyl-phosphatidylinositol membrane anchor of

RT the leishmania major promastigote surface protease.";
RL J. Biol. Chem. 265:16955-16964(1990).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).
RX MEDLINE=95406217; PubMed=7675788;
RA Schlegelhauf E., Etges R., Metcalf P.;
RT "Crystallization and preliminary X-ray diffraction studies of
RT leishmanolysin, the major surface metalloprotease from Leishmania
RT major.";
RL Proteins 22:58-66(1995).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.86 ANGSTROMS).
RX MEDLINE=98416698; PubMed=9793094;
RA Schlegelhauf E., Etges R., Metcalf P.;
RT "The crystal structure of the Leishmania major surface protease
RT leishmanolysin.";
RL Structure 6:1035-1046(1998).
CC -!- FUNCTION: Has an integral role during the infection of macrophages
CC in the mammalian host.
CC -!- CATALYTIC ACTIVITY: Preference for hydrophobic residues at p1 and
CC p1' and basic residues at p2 and p3'. A model nonapeptide is
CC cleaved at -Ala-Tyr-[Leu-Lys-Lys-
CC -!- COFACTOR: Binds 1 zinc ion per subunit.
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -!- PTM: THE PHOSPHATIDYLINOSITOL MOIETY OF THE GPI-ANCHOR CONTAINS A
CC FULLY SATURATED, UNBRANCHED 1-O-ALKYL CHAIN (MAINLY C24:0) AND A
CC MIXTURE OF FULLY SATURATED UNBRANCHED 2-O-ACYL CHAINS (C12:0,
CC C14:0, C16:0, AND C18:0).
CC -!- SIMILARITY: Belongs to peptidase family M8.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Y00647; CAA68673.1; -.
DR PIR; PLO221; PLO221.
DR PDB; 1LMJ; 17-SEP-97.
DR MEROPS; M08.001; -.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR InterPro; IPR001577; Peptidase_M8.
DR Pfam; PF01457; Peptidase_M8; 1.
DR PRINTS; PR00782; LSHMANOLYSIN.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydroxylase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;
KW Zymogen; Signal; Cell adhesion; GPI-anchor; 3D-structure; Lipoprotein.
FT SIGNAL 1 39
FT PROPEP 40 100
FT CHAIN 101 577
FT PROPEP 578 602
FT METAL 264 264
FT ACT SITE 265 265
FT METAL 268 268
FT METAL 334 334
FT DISULFID 125 142
FT DISULFID 191 230
FT DISULFID 314 386
FT DISULFID 393 455
FT DISULFID 406 425
FT DISULFID 415 489
FT DISULFID 466 510
FT DISULFID 515 565
FT DISULFID 535 558
FT CARBOHYD 300 300
FT CARBOHYD 407 407
FT LIPID 577 577
FT STRAND 101 102
FT STRAND 107 108
FT STRAND 111 114
FT HELIX 116 119
N-LINKED (GLCNAC. . .) (POTENTIAL).
N-LINKED (GLCNAC. . .) (POTENTIAL).
GPI-anchor amidated asparagine.

FT TURN 121 122
FT TURN 128 129
FT STRAND 131 133
FT STRAND 139 141
FT HELIX 144 146
FT HELIX 150 158
FT TURN 159 159
FT HELIX 160 169
FT TURN 170 171
FT STRAND 172 174
FT STRAND 177 178
FT STRAND 180 181
FT TURN 189 190
FT HELIX 191 193
FT HELIX 198 202
FT TURN 203 203
FT STRAND 205 206
FT STRAND 210 215
FT TURN 221 222
FT STRAND 226 232
FT TURN 234 235
FT STRAND 238 244
FT HELIX 247 249
FT HELIX 256 269
FT TURN 270 271
FT HELIX 274 279
FT TURN 280 281
FT STRAND 283 286
FT HELIX 289 291
FT STRAND 296 299
FT HELIX 302 312
FT TURN 313 313
FT TURN 315 316
FT STRAND 320 322
FT TURN 328 332
FT STRAND 335 335
FT TURN 337 339
FT TURN 341 342
FT STRAND 343 343
FT TURN 344 345
FT STRAND 353 353
FT HELIX 356 364
FT TURN 365 366
FT STRAND 369 370
FT HELIX 372 374
FT TURN 380 383
FT HELIX 386 390
FT STRAND 394 395
FT TURN 396 397
FT STRAND 398 399
FT TURN 402 404
FT STRAND 413 414
FT TURN 417 418
FT STRAND 421 425
FT STRAND 428 429
FT HELIX 435 437
FT TURN 443 444
FT STRAND 445 446
FT TURN 450 454
FT STRAND 458 465
FT TURN 466 467
FT HELIX 470 472
FT TURN 475 477
FT HELIX 480 480
FT TURN 485 486
FT STRAND 487 494
FT STRAND 496 496
FT TURN 506 516
FT STRAND 517 520
FT TURN 521 525
FT STRAND 527 528
FT STRAND 533 534
FT TURN 537 538

FT STRAND 540 542
 FT HELIX 543 545
 FT TURN 546 546
 FT STRAND 550 550
 FT TURN 552 553
 FT STRAND 555 557
 FT HELIX 561 565
 FT TURN 566 567
 FT HELIX 569 572
 FT TURN 573 573
 SQ SEQUENCE 602 AA; 63953 MW; 982EF3245D87C43E CRC64;

Query Match 11.2%; Score 87; DB 1; Length 602;
 Best Local Similarity 89.5%; Pred. No. 4.8;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 85 ADNYDQLVTRVVTHEMAHA 103
 Db 251 ASRYDQLVTRVVTHEMAHA 269

Search completed: March 11, 2004, 18:34:57
 Job time : 6.3 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:23:54 ; Search time 30.5 Seconds
(without alignment)
1562.074 Million cell updates/sec

Title: US-09-543-407-22

Perfect score: 776
Sequence: 1 MLLKVAAPAAIVSGSALA.....DSSVMVROVFGNNATANY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPREMBL_25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriophage:*
- 17: sp_archaeal:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	673	86.7	152	2	033802
2	566.5	73.0	150	2	07X243
3	538	69.3	149	2	07X240
4	496.5	64.0	152	16	08CW63
5	421.5	54.3	150	2	07X237
6	305.5	39.4	76	2	054069
7	122	15.7	29	2	09S3J5
8	115	14.8	139	16	08B1H3
9	110	14.2	130	16	08B1H4
10	109.5	14.1	502	16	08B1H4
11	109	14.0	1748	5	094821
12	109	14.0	2174	16	092U08
13	106.5	13.7	3501	16	08Y106
14	106.5	13.7	3552	16	08XSD6
15	106	13.7	179	2	033801
16	105	13.5	1422	16	08B1H3

17	104	13.4	714	16	Q7USX6
18	102	13.1	191	10	Q7XDR3
19	102	13.1	586	5	Q9NAJ4
20	98	12.6	614	16	Q7YR8
21	98	12.6	615	16	P95249
22	96.5	12.4	171	16	08GJ13
23	96	12.4	2310	5	Q9W2U7
24	95.5	12.3	3659	16	Q9BLN6
25	95	12.2	589	16	Q7TX53
26	95	12.2	590	16	053309
27	95	12.2	603	16	08V766
28	94.5	12.2	157	16	08BHC0
29	94.5	12.2	738	5	002402
30	94.5	12.2	7716	16	Q7UWZ8
31	94	12.1	407	5	Q20151
32	93.5	12.0	145	16	Q8U6N9
33	93.5	12.0	645	16	Q7U1C5
34	93.5	12.0	646	16	053818
35	93.5	12.0	1286	2	Q841Y5
36	93.5	12.0	1333	16	Q8PD38
37	92.5	11.9	151	16	Q7UC21
38	92.5	11.9	154	16	08GJ15
39	92.5	11.9	160	16	08CW64
40	92.5	11.9	160	16	Q83RU7
41	92.5	11.9	453	5	Q9NGF7
42	92.5	11.9	453	5	Q9N6M8
43	92	11.9	196	10	Q22638
44	91.5	11.8	151	2	Q7X238
45	91.5	11.8	453	5	Q9NGF6

ALIGNMENTS

RESULT 1	
033802	PRELIMINARY; PRT; 152 AA.
ID	033802
AC	033802;
DT	01-JAN-1998 (TREMBLrel. 05, Created)
DT	01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE	Agfa protein (Fragment).
GN	Agfa.
OS	Salmonella typhimurium.
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC	Enterobacteriaceae; Salmonella.
OX	NCBI_TaxID=602;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=98053981; PubMed=9393832;
RA	Sukhopolvi S.S., Lorentz R.G., Gordon J.I., Bian Z., Pfeiffer J.D.,
RA	Normark S.J., Rhen M.;
RT	"Expression of thin, aggregative fimbriae promotes interaction of
RT	RT Salmonella typhimurium SR-11 with mouse small intestinal epithelial
RT	cells";
RL	Infect. Immun. 65:5320-5325(1997).
DR	EMBL; AJ000514; CA04151.1; -
FT	NON TER 152 152
SO	SEQUENCE 152 AA; 15401 MW; 9DA7DADC2364B006 CRC64;
Query Match	86.7%; Score 673; DB 2; Length 152;
Best Local Similarity	89.4%; Pred. No. 3.7e-47;
Matches 135; Conservative	3; Mismatches 13; Indels 0; Gaps 0;
QY	1 MLLKVAAPAAIVSGSALAGVPPWGGGNGGNGSGSPSTSIYQGSANALALQ 60
DB	1 MLLKVAAPAAIVSGSAVAVPPWGGGNGGNGSGSPSTSIYQGSANALALQ 60
QY	61 SPARKSETTITSGYGNGADVGQAGADNYDQTLTRVVTTHMAHADQNNAKSDITVQYGG 120
DB	61 SPARKSETTITSGYGNGADVGQAGADNSTRITETLQNGFRNNAITIDQNNAKSDITVQYGG 120
QY	121 NNALVNOTASDSSVMVROVFGNNATANY 151

Db 121 NNALVNOTASDSSVWVROVGFNNATPANOY 151

RESULT 2

07X243 PRELIMINARY; PRT; 150 AA.
AC 07X243;
DT 01-OCT-2003 (TREMBLrel. 25, Created)
DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)
DE Curlin-csgA protein.
GN CSGA.
OS Citrobacter sp. Fec2.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Citrobacter.
OX NCBI_Taxid=213763;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fec2;
RA Zogaj X., Bokranz W., Nimtz M., Romling U.;
RT "Production of Cellulose and Curli Fimbriae by Members of the Family Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158(2003).
DR EMBL; AJ515700; CAD56672.1; -.
SQ SEQUENCE 150 AA; 15016 MW; 1D7141BD6973DC6 CRC64;

Query Match 73.0%; Score 566.5; DB 2; Length 150;
Best Local Similarity 77.5%; Pred. No. 1.6e-38;
Matches 117; Conservative 11; Mismatches 22; Indels 1; Gaps 1;

Qy 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNNHGGSGPDSLTSTIYOGSANAALAQ 60
Db 1 MKLLQVAAFAAIVVSGSALAGVVPQWGGGCGG-GGGSSSGESTLSTIYOGSANAALAQ 59
Qy 61 SDARKSETTTTQSGYNGADVGQGDNDYDQVTVFVTHEMAHADQWNAKNSDITVQYGG 120
Db 60 SDARKSDTTTHQNGFGNADVGQGSNDSTIDLTQNGFGNNATIDQWNGKNSDITVQYGG 119
Qy 121 NNALVNOTASDSSVWVROVGFNNATPANOY 151
Db 120 NNALVNOTASDSSVWVROVGFNNATPANOY 150

RESULT 3

07X240 PRELIMINARY; PRT; 149 AA.
AC 07X240;
DT 01-OCT-2003 (TREMBLrel. 25, Created)
DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Curlin-csgA protein.
GN CSGA.
OS Citrobacter freundii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Citrobacter.
OX NCBI_Taxid=546;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fec4;
RA Zogaj X., Bokranz W., Nimtz M., Romling U.;
RT "Production of Cellulose and Curli Fimbriae by Members of the Family Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158(2003).
DR EMBL; AJ515701; CAD56675.1; -.
SQ SEQUENCE 149 AA; 15260 MW; 946DD52017F648FD CRC64;

Query Match 69.3%; Score 538; DB 2; Length 149;
Best Local Similarity 72.2%; Pred. No. 3.2e-36;
Matches 109; Conservative 17; Mismatches 23; Indels 2; Gaps 1;

Qy 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNNHGGSGPDSLTSTIYOGSANAALAQ 60

Db 1 MKLLKVAFAAIVVSGSALAGVVPQW--GGNHGGSGNYGPDSSLTSTIYOGSNNANALQ 58

Qy 61 SDARKSETTTTQSGYNGADVGQGDNDYDQVTVFVTHEMAHADQWNAKNSDITVQYGG 120
Db 59 SDARKSDTTTHQNGFGNADVGQGSNDSTIDLTQNGFGNNATIDQWNGKNSDITVQYGG 118

Qy 121 NNALVNOTASDSSVWVROVGFNNATPANOY 151
Db 119 NNALVNOTASDSSVWVROVGFNNATPANOY 149

RESULT 4

08CW63 PRELIMINARY; PRT; 152 AA.
AC 08CW63;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DE Major curlin subunit precursor.
GN CSGA OR C1306.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_Taxid=217992;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O6:HT / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G., III, Redford P., Roesch P., Raske D., Buckles E.L., Lion S.-R., Boutin A., Hackett J., Stroud D., Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T., Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
DR EMBL; AE016759; AAN79779.1; -.
KM Complete proteome.
SQ SEQUENCE 152 AA; 15064 MW; 3BA57F34C1240E83 CRC64;

Query Match 64.0%; Score 496.5; DB 16; Length 152;
Best Local Similarity 67.8%; Pred. No. 7.7e-33;
Matches 103; Conservative 17; Mismatches 31; Indels 1; Gaps 1;

Qy 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNNHGGSGPDSLTSTIYOGSANAALAQ 59
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNNHGGSGPDSLTSTIYOGSANAALAQ 60
Qy 60 QSDARKSETTTTQSGYNGADVGQGDNDYDQVTVFVTHEMAHADQWNAKNSDITVQYGG 119
Db 61 QADARKSDTTTHQNGFGNADVGQGSNDSTIDLTQNGFGNNATIDQWNGKNSDITVQYGG 120
Qy 120 NNALVNOTASDSSVWVROVGFNNATPANOY 151
Db 121 NNALVNOTASDSSVWVROVGFNNATPANOY 152

RESULT 5

07X237 PRELIMINARY; PRT; 150 AA.
AC 07X237;
DT 01-OCT-2003 (TREMBLrel. 25, Created)
DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Curlin-csgA protein.
GN CSGA.
OS Enterobacter sakazakii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Enterobacter.
OX NCBI_Taxid=28141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fec39;
RA Zogaj X., Bokranz W., Nimtz M., Romling U.;


```

RT "Production of Cellulose and Curli Fimbriae by Members of the Family
  Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158 (2003).
SQ SEQUENCE 150 AA; 15112 MW; 5D8B2D872DF15F3 CRC64;

Query Match      54.3%; Score 421.5; DB 2; Length 150;
Best Local Similarity 58.9%; Pred. No. 9.2e-27;
Matches 89; Conservative 25; Mismatches 36; Indels 1; Gaps 1;

QY 1 MKLTKVAAPAAIVSGSALAGVVPQWGGGNGGSGPSTLSIYOGSANAALALD 60
  DB 1 MKRIFAAALAAIVSGSANAAMINQ-GGKHGHGHGGYGGPSTLTINYGGSALALD 59
  QY 61 SDARKSETTITGSGYGAGDVGGADNYDQLVTRVVTHEMAADQWNAKNSDITVGG 120
  DB 60 TDARNSVILNISQGGGAGDVGGSDSSINLTONGGNSATLDQWNSKDSVNVVGGYGG 119
  QY 121 NNAALVNOTASDSSVNVQVGFNNATANOY 151
  DB 120 LNALVDTQASNSTVNTQIGFNHATAHQY 150

RESULT 6
ID 054069 PRELIMINARY; PRT; 76 AA.
AC 054069;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DE 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
GN SEF17 fimbriin (Fragment).
OS Salmonella enteritidis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
  Enterobacteriaceae; Salmonella.
OX NCBI_Taxid=592;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SE30;
RA Cox J.M., Egleros S., Woolcock J.B.;
RT "Influence of Salmonella enteritidis in chickens correlates with
  colony morphology and expression of SEF17 fimbriae.";
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U53207; AAA98671.1; -.
FT NON TER 1 1
FT NON TER 76 76
SQ SEQUENCE 76 AA; 7704 MW; 2FDS411241A7BCB1 CRC64;

Query Match      39.4%; Score 305.5; DB 2; Length 76;
Best Local Similarity 54.5%; Pred. No. 1e-17;
Matches 66; Conservative 4; Mismatches 6; Indels 45; Gaps 1;

QY 30 GHNNGGSSGPGSTLSIYOGSANAALALQSDARKSETTITGSGYGAGDVGGADNYD 89
  DB 1 GHNXGGSSGPGSTLSIYOGSANAALALQSDARKSETTITGSGYGAGDVGGADN-- 58
  QY 90 QLVTRVVTHEMAADQWNAKNSDITVQYGGNNAALVNTASDSSVNVQVGFNNATAN 149
  DB 59 -----STIELTONGFRNNATID 75
  QY 150 Q 150
  DB 76 Q 76

RESULT 7
ID 095335 PRELIMINARY; PRT; 29 AA.
AC 095335;
DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (T-EMBLrel. 16, Last annotation update)
DE Curlin subunit monomer (Fragment).

```

```

GN CSGA.
OC Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
  Enterobacteriaceae; Escherichia.
OX NCBI_Taxid=562;
RN [1]
RP SEQUENCE FROM N.A.
RC TRANSPOSON-insertion sequence IS1;
RX MEDLINE=99314153; PubMed=10386375;
RA La Ragione R.M., Collighan R.J., Woodward M.J.;
RT "Non-curliation of Escherichia coli O78:X80 isolates associated with
  IS1 inserti on in cs8b and reduced persistence in poultry infection.";
RL FEMS Microbiol. Lett. 175:247-253 (1999).
DR EMBL; AJ131756; CAB45380.1; -.
FT NON TER 29 29
FT NON TER 29 29
SQ SEQUENCE 29 AA; 2789 MW; E290DEC07ABBB243 CRC64;

Query Match      15.7%; Score 122; DB 2; Length 29;
Best Local Similarity 89.7%; Pred. No. 0.0025;
Matches 26; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKLTKVAAPAAIVSGSALAGVVPQWGGG 29
  DB 1 MKLTKVAALAAIVSGSALAGVVPQYGGG 29

RESULT 8
ID 08EIH3 PRELIMINARY; PRT; 139 AA.
AC 08EIH3;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DE 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
GN Minor curlin subunit CsgB, putative.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
  Alteromonadaceae; Shewanella.
OX NCBI_Taxid=70863;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NR-1;
RX MEDLINE=22297686; PubMed=12368813;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
  Read T.D., Eisen J.A., Seshadri R., Ward N., Meche B., Clayton R.A.,
  Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
  DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolony J.F.,
  Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
  Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,
  Mueller J., Kouri H., Gill J., Uetreback T.R., McDonald L.A.,
  Feldebyum T.V., Smith H.O., Venter J.C., Neilson K.H., Fraser C.M.;
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
  Shewanella oneidensis.";
RL Nat. Biotechnol. 20:1118-1123 (2002).
DR EMBL; AE015532; AAN53942.1; -.
KW Complete proteome.
SQ SEQUENCE 139 AA; 14811 MW; 41B1C1FA76957920 CRC64;

Query Match      14.8%; Score 115; DB 16; Length 139;
Best Local Similarity 28.3%; Pred. No. 0.061;
Matches 30; Conservative 19; Mismatches 33; Indels 24; Gaps 2;

QY 39 SGPSTLSIYOGSANAALALQSDARKSETTITGSGYGAGDVGGADNYDQLVTRVVT 98
  DB 41 SGRDNLIDLVQGTANQGI-----VFQSGSDSSAYVTQAGNDNISLVLTQIGT- 87
  QY 99 EMAHADQWNAKNSDITVQYGGNNAALVNTASDSSVNVQVGFNN 144
  DB 88 -----NNEVQLLVQGAQNKASITQIGNDNLVQLNGSGSN 122

RESULT 9

```

```

Q89J14
ID 089J14 PRELIMINARY; PRT; 130 AA.
AC 089J14;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DE B115299 protein.
GN B115299.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobiium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RX MEDLINE=22484998; PubMed=12597275;
RA Kaneo T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA Sasamoto S., Matsumoto A., Idesawa K., Irituchi M., Kawashima K.,
RA Kohata M., Matsumoto M., Shimo S., Tsuruoka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110."
RL DNA Res. 9:189-197 (2002).
DR EMBL; AP005954; BAC50564.1; -.
KW Complete proteome.
SQ SEQUENCE 130 AA; 12699 MW; ACFB2D6A48D260F CRC64;

Query Match 14.2%; Score 110; DB 16; Length 130;
Best Local Similarity 25.8%; Pred. No. 0.14;
Matches 39; Conservative 28; Mismatches 56; Indels 28; Gaps 5;

QY 4 LKVAAPALVSGSALAGVPMQGGGNNHGGSSGPDSTLSIYQGSANALALQSDA 63
DB 1 KRITLVATVATLALSLATTVDAQ-----AGNSA-----SVLQFETITSSPISQTS 45
QY 64 RSEETITQSGYGADVDGQGANVDQVLTVRVTHMAHADQW-----NAKNSDITVQYG 119
DB 46 TSNNAITTIQFGATNTATTLQTS-----LTVNTAVTQGGTTATASNTALT-QQVG 96
QY 120 GNNALVNOQTASDSSVWVRQVFGNNATANQ 150
DB 97 GNSSLIGQIGANNVTAGVGQGIINGSTILQ 127

RESULT 10
Q8EIH4 PRELIMINARY; PRT; 502 AA.
ID 08EIH4;
AC 08EIH4;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DE 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Conserved hypothetical protein.
GN S00865.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.
OX NCBI_TaxID=70863;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MR-1;
RX MEDLINE=22297866; PubMed=12368813;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Bisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
RA Meyer T., Tsapin A., Scott J., Beaman M., Birkhac L., Daugherty S.,
RA DeBoy R.T., Dodson R.U., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Imprim L.A., White O., Wolf A.M.,
RA Vamathevan J., Weidman J., Imprial M., Lee K., Berry K., Lee C.,
RA Fiedlyum T.V., Smith H.O., Venter J.C., Neeson K.H., Fraser C.M.;
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis."
RL Nat. Biotechnol. 20:1118-1123 (2002).
DR EMBL; AB015532; AAN53941.1; -.

```

```

DR TIGR; S00865; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 502 AA; 52441 MW; D08CA23D6C46B62D CRC64;

Query Match 14.1%; Score 109.5; DB 16; Length 502;
Best Local Similarity 23.8%; Pred. No. 0.79;
Matches 39; Conservative 26; Mismatches 50; Indels 49; Gaps 5;

QY 29 GGNHGGSSGPDSTLSIYQGSANAA-----LALQSDARKSETTITQSGYNG 78
DB 347 GDNNELVAFATGFRDNGSIEISQEGDANFAYVDATGNDNEVNEVDGQNETITTVBG-NNN 405
QY 79 ADVG-----QGADNDQVLTVRVTHMAHADQW-----N 107
DB 406 ADVTALQHRGDLNLIDLIIEGDENAEIT-----QASGMMVGSDSSSFAASSFGVS 458
QY 108 AKNSDITVQGYGNNALVNOQTASDSSVWVRQVFGNNATANQY 151
DB 459 GDNNSLMTITQGTGNDNLVLSGQAGNNNSISVTSGDNNVATVVOY 502

RESULT 11
Q94821 PRELIMINARY; PRT; 1748 AA.
ID 094821;
AC 094821; P92146; P92145; P92144; P92143; P92142; P92141; Q94820;
DT 01-FEB-1997 (TREMBlrel. 02, Created)
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE CNUB protein.
GN CNUB.
OS Tetrahymena thermophila.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
OC Tetrahymenina; Tetrahymena.
OX NCBI_TaxID=5911;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88189811; PubMed=3357771;
RA Martindale D.W., Taylor F.M.;
RT "Multiple introns in a conjugation-specific gene from Tetrahymena
RT thermophila."
RL Nucleic Acids Res. 16:2189-2201 (1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94051569; PubMed=8233798;
RA Taylor F.M., Martindale D.W.;
RT "Retroviral-type zinc fingers and glycine-rich repeats in a protein
RT encoded by cnuB, a Tetrahymena gene active during meiosis."
RL Nucleic Acids Res. 21:4610-4614 (1993).
DR EMBL; X06462; CAB37323.1; -.
DR EMBL; L03710; AAC37171.1; -.
DR PIR; S42136; S42136.
DR HSP; P05888; IAAF.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF00098; Zf_CCHC; 7.
DR PRINTS; PR00939; C2HCZNFINGER.
DR SMART; SM00343; Znf_CCHC; 7.
DR PROSITE; PS50158; Zf_CCHC; 7.
FT CONFLICT 251 251 M->I (IN REF. 1).
FT FT 256 256 I->N (IN REF. 1).
SQ SEQUENCE 1748 AA; 199624 MW; 0B03F210104008A3 CRC64;

Query Match 14.0%; Score 109; DB 5; Length 1748;
Best Local Similarity 33.6%; Pred. No. 3.9;
Matches 37; Conservative 11; Mismatches 32; Indels 30; Gaps 6;

QY 25 QMGCGGNHNGG---GNSGPDSTLSIYQGSANALALQSDARKSETTIT---QSGYGV 77
DB 1440 QFGGGGNSNGQSGWGTSSGSDWN-----CSNVGSESTTSSGCGWSSGSGN 1685
QY 78 GADVGGGA-DNYDQVLTVRVTHMAHADQWNAKNSDITVGO--YGGNNA 124
DB 1686 QTGGGWSNDNQQD-----QNTNTGGGCGSSNSQNTNNSSWGSNNQA 1729

```

```

RESULT 12
Q92U08      PRELIMINARY;      PRT; 2174 AA.
AC  Q92U08;
DT  01-DEC-2001 (TrEMBLrel. 19, Created)
DT  01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE  Hypothetical glycine-rich protein SMD21548.
GN  RB0989 OR SMD21548.
OS  Rhizobium meliloti (Sinorhizobium meliloti).
OC  Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC  Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX  NCBI_TaxID=382;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=1021;
RX  MEDLINE=21396508; PubMed=11481431;
RA  Finan T.M., Weidner S., Wong K., Buhmester J., Chain P.,
RA  Vorheeler F.J., Hernandez-Lucas I., Becker A., Gouzy J.,
RA  Golding B., Puehler A.;
RT  "The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-
RT  fixing endosymbiont Sinorhizobium meliloti."
RL  Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).
DR  EMBL; AL603645; CAC49389.1; -.
DR  PIR; E95965; E95965.
DR  GO; GO:0046821; C:extrachromosomal DNA; IEA.
DR  GO; GO:0004601; P:peroxidase activity; IEA.
DR  GO; GO:0006979; P:response to oxidative stress; IEA.
DR  InterPro; IPR005546; Autotransporter.
DR  InterPro; IPR002016; Peroxidase.
DR  Pfam; PF03797; Autotransporter; 1.
DR  PROSITE; PS00435; PEROXIDASE_1; 1.
DR  PROSITE; PS00583; PFKS_KINASES_1; 2.
KW  Plasmid; Hypothetical protein; Complete proteome.
SQ  SEQUENCE 2174 AA; 203314 MW; 008E868297B44182 CRC64;

Query Match      14.0%; Score 109; DB 16; Length 2174;
Best Local Similarity 27.0%; Pred. No. 5.1;
Matches 40; Conservative 20; Mismatches 52; Indels 36; Gaps 7;

QY  11 AIVVSGSALAGVVPQ--WGGGNNHNGCGNSGPDSTLSIYQGS-----ANAA----- 56
DB  693 AITATAGAVGILAQSIGGGCGN---CGNATGCDAGCGFQIGGGCGGYANTAVNGFK 749
QY  57 -IALQSDARKSETTTTQSGYNGADVQCGADNYDLVTRVVTHEMAHADQMAKNSDITY 115
DB  750 GLTLTTQGSHPAGIVAQS-VGGGGGTGTASYSAGI-----GFTASVAV 793
QY  116 GOYGGNNA--LYNOTASDSVVVRQYG 141
DB  794 GGTGGNGAGGEVSVLTDASIRTTGGG 821

RESULT 13
Q8Y106      PRELIMINARY;      PRT; 3501 AA.
AC  Q8Y106;
DT  01-MAR-2002 (TrEMBLrel. 20, Created)
DT  01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DE  Probable hemagglutinin-related protein.
GN  RSC0887 OR RS06116.
OS  Ralstonia solanacearum (Pseudomonas solanacearum).
OC  Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC  Burkholderiaceae; Ralstonia.
OX  NCBI_TaxID=305;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=GMI1000;

```

```

RX  MEDLINE=21681879; PubMed=11823852;
RA  Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA  Arlat M., Billault A., Brottier P., Camus J.C., Catolico L.,
RA  Chandler M., Choiane N., Claudel-Renard C., Cunac S., Demange N.,
RA  Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
RA  Signier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA  Weisenbach J., Boucher C.A.;
RT  "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL  Nature 415:497-502(2002).
DR  EMBL; AL646061; CAD14589.1; -.
DR  GO; GO:0004519; F:endonuclease activity; IEA.
DR  GO; GO:0003676; F:nucleic acid binding; IEA.
DR  InterPro; IPR001604; Endonuclease.
DR  InterPro; IPR008619; F11_haemaggl.
DR  InterPro; IPR008638; Haemaggl. act.
DR  Pfam; PF05594; F11_haemaggl; 20.
DR  Pfam; PF05860; Haemaggl act; 1.
DR  PROSITE; PS01070; NUCLEASE_NON_SPEC; 1.
KW  Complete proteome.
SQ  SEQUENCE 3501 AA; 348421 MW; 290B41C99018A107 CRC64;

Query Match      13.7%; Score 106.5; DB 16; Length 3501;
Best Local Similarity 29.5%; Pred. No. 14;
Matches 43; Conservative 19; Mismatches 45; Indels 39; Gaps 7;

QY  15 SGGSALAGVVPQWGGGNNHNGG--NSSGPDSTLSIYQGSANNA----- 56
DB  2431 SGGSHFTAGPSPWDLGRNVGGGPNSSG---VGLARYGSAHSDNNAAGNSSRQNASVIG 2486
QY  57 LIALQSDARKSETTTTQSGYNGADVQ--CGADNYDLVTRVVTHEMA--HADQMAK 110
DB  2487 KSVQVAFRTGDTIVSSGSGISALSDVDLAKQK-----VDIVAGNDTSRRHD-----H 2535
QY  111 SDITVGYGNNAAALVNOTASDSVVM 136
DB  2536 SDRTTDLGNGYSGTVGRASSTL 2561

RESULT 14
Q8XSD6      PRELIMINARY;      PRT; 3552 AA.
AC  Q8XSD6;
DT  01-MAR-2002 (TrEMBLrel. 20, Created)
DT  01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DE  Probable hemagglutinin-related protein.
GN  RSP0540 OR RS06117.
OS  Ralstonia solanacearum (Pseudomonas solanacearum).
OC  Plasmid megaplasmid.
OC  Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC  Burkholderiaceae; Ralstonia.
OX  NCBI_TaxID=305;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=GMI1000;
RX  MEDLINE=21681879; PubMed=11823852;
RA  Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA  Arlat M., Billault A., Brottier P., Camus J.C., Catolico L.,
RA  Chandler M., Choiane N., Claudel-Renard C., Cunac S., Demange N.,
RA  Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
RA  Signier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA  Weisenbach J., Boucher C.A.;
RT  "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL  Nature 415:497-502(2002).
DR  EMBL; AL646079; CAD17691.1; -.
DR  GO; GO:0046821; C:extrachromosomal DNA; IEA.
DR  GO; GO:0004519; F:endonuclease activity; IEA.
DR  GO; GO:0003676; F:nucleic acid binding; IEA.
DR  InterPro; IPR001604; Endonuclease.
DR  InterPro; IPR008619; F11_haemaggl.
DR  InterPro; IPR008638; Haemaggl act.
DR  Pfam; PF05594; F11_haemaggl; 20.
DR  Pfam; PF05860; Haemaggl_act; 1.

```

DR PROSITE; PS01070; NUCLEASE NON_SPEC; 1.
 KQ Plasmid; Complete proteome.
 SQ SEQUENCE 3552 AA; 35293 AA; CS432AABE2CCF59C CRC64;

Query Match 13.7%; Score 106.5; DB 16; Length 3552;
 Best Local Similarity 29.5%; Pred. No. 15;
 Matches 43; Conservative 19; Mismatches 45; Indels 39; Gaps 7;

QY 15 GSAALAGVVPQWGGGNNHGGG-NSSGPDSTLSIYQGSANAA----- 56
 DB 2430 GSHSTGTPSGWGDIGRVVGGGPNSSG---VGLAPYGAHSDADNAAGNSSRQNASVTVG 2485
 QY 57 LALQSDARKSETTITQSGYGNAGADV---QGANVDQLVTRVYTHEMA--HADQWNAKN 110
 DB 2486 KSVQVQARTGDTIVSGSGISALSDVDLAKQK-----VDIVAGNDTSSRHED-----H 2534
 QY 111 SDITVGYGNNALVNOTASDSSVM 136
 DB 2535 SDRITIGDLGGNGSGTVGRSASTL 2560

RESULT 15

033801 PRELIMINARY; PRT; 179 AA.
 AC 033801;
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE AgfB protein.
 GN AGFB.
 OS Saimonella typhimurium.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Saimonella.
 OX NCBI_TaxID=602;
 RN [1]
 RP SEQUENCE FROM N.A.
 FX MEDLINE=98053981; PubMed=9393832;
 RA Sukupolvi S.S., Lorentz R.G., Gordon J.I., Bian Z., Pfeiffer J.D.,
 Normark S.J., Rhen M.;
 RT "Expression of thin, aggregative fimbriae promotes interaction of
 RT Saimonella typhimurium SR-11 with mouse small intestinal epithelial
 RT cells.";
 RL Infect. Immun. 65:5320-5325(1997).
 DR EMBL; AJ000514; CAA04150.1; -
 SQ SEQUENCE 179 AA; 19318 MM; A2BCCB648B3C0B0B CRC64;

Query Match 13.7%; Score 106; DB 2; Length 179;
 Best Local Similarity 32.3%; Pred. No. 0.44;
 Matches 41; Conservative 17; Mismatches 55; Indels 14; Gaps 5;

QY 16 GSALAGVVPQWGGGNNHGGGSSGPDSTLSIYQGSANALAIQS--DARKSETTITQ 72
 DB 63 GSKLSTVLSQ-ERGGNRAKVDQGNVNFATIEGTGNANDASISQSAVNSAISAIIQ 121
 QY 73 SGYGNAGADVQGANVYDQLVTRVYTHEMAHADQWNAKNSDITVGYGNNALVNOTASD 132
 DB 122 KGSNKNKANIQ---YGTOKTAVVQKOSH---AIOANIT--QYGTOKTAVVQKOSH 171
 QY 133 SSVWVRQ 139
 DB 172 MAIRVTQ 178

Search completed: March 11, 2004, 18:40:43
 Job time : 31.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:13:53 ; Search time 45.9 Seconds
(without alignments)
929.514 Million cell updates/sec

Title: US-09-543-407-24

Perfect score: 775
Sequence: 1 MLKLVAAFAAIVSGSALA.....DSSVMVROYFGNNATANY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: A_Geneseq_293a04:*

1: geneseqp19808:*\n2: geneseqp19908:*\n3: geneseqp20008:*\n4: geneseqp20018:*\n5: geneseqp20028:*\n6: geneseqp20038:*\n7: geneseqp20038:*\n8: geneseqp20048:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	775	100.0	151	3	AAB36352
2	714	92.1	151	3	AAB36347
3	700	90.3	151	3	AAB36355
4	690	89.0	151	2	AAR74625
5	689	89.0	151	3	AAB36341
6	689	88.9	151	2	AAR73570
7	655	84.5	151	3	AAB36346
8	609	78.6	151	3	AAB36353
9	609	78.6	151	3	AAB36351
10	608	78.5	151	3	AAB36349
11	606	78.2	151	3	AAB36350
12	600	77.4	151	3	AAB36354
13	575	74.2	151	3	AAB36348
14	533	68.8	151	3	AAB36343
15	528	68.1	151	7	ABR82651
16	505	65.2	120	2	AAR62761
17	505	65.2	120	2	AAR73559
18	455	58.7	142	2	AAR32664
19	383	49.4	122	2	AAR32663
20	237	30.6	45	3	AAB36316
21	132	17.0	22	3	AAB36318
22	123	15.9	23	3	AAB36331
23	123	15.9	23	3	AAB36336
24	123	15.9	23	3	AAB36338
25	113	14.8	22	3	AAB36325

26	115	14.8	22	3	AAB36339	AAB36339 Salmonella
27	115	14.8	22	3	AAB36320	AAB36320 Salmonella
28	113	14.6	24	7	ABR82644	ABR82644 E. coli C
29	109	14.1	23	3	AAB36340	AAB36340 Salmonella
30	109	14.1	23	3	AAB36324	AAB36324 Salmonella
31	109	14.1	23	3	AAB36319	AAB36319 Salmonella
32	98.5	12.7	151	3	AAB36342	AAB36342 Salmonella
33	98	12.6	26	7	ABR82649	ABR82649 E. coli V
34	98	12.6	26	7	ABR82645	ABR82645 E. coli C
35	98	12.6	903	2	AAR35006	AAR35006 POLYANGU
36	96.5	12.5	151	3	AAB36344	AAB36344 Escherichia
37	96	12.4	19	3	AAB36323	AAB36323 Salmonella
38	96	12.4	19	3	AAB36336	AAB36336 Salmonella
39	96	12.4	19	3	AAB36328	AAB36328 Salmonella
40	92.5	11.9	850	4	ABR85764	ABR85764 Drosophila
41	92.5	11.9	1028	4	ABR82708	ABR82708 Drosophila
42	92	11.9	23	3	AAB36331	AAB36331 Escherichia
43	91	11.7	688	5	ABP74039	ABP74039 Candida
44	90.5	11.7	378	4	ABR86461	ABR86461 Drosophila
45	90.5	11.7	520	6	AAO16497	AAO16497 Argiophe

ALIGNMENTS

RESULT 1	AAB36352	
ID	AAB36352 standard; protein; 151 AA.	
XX		
AC	AAB36352;	
XX		
DT	26-FEB-2001 (first entry)	
XX		
DE	Agfa::PT3#7 amino acid sequence SEQ ID NO:24.	
XX		
KW	Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;	
XX	vaccine; immune response; immunogen.	
OS	Salmonella enteritidis.	
OS	Escherichia coli.	
XX	Synthetic.	
PN	WO200060102-A2.	
XX		
PD	12-OCT-2000.	
XX		
PF	05-APR-2000; 2000WO-CA00035C.	
XX		
PR	05-APR-1999; 99US-0127888P.	
XX		
PA	(UUYV-) UNIV VICTORIA.	
XX		
PI	White AP, Doran JL, Collison SK, Kay WW;	
XX		
DR	WPI; 2000-672631/65.	
XX		
NR	N-PSDB; AAC64628.	
PT	Recombinant agfa gene having a segment replaced by a foreign DNA sequence	
XX	which encodes foreign epitope or antigen, expresses recombinant Agfa	
PT	protein useful for eliciting immune response in animal.	
XX		
PS	Disclosure; Page 138; 139pp; English.	
XX		
CC	The present invention describes a recombinant agfa gene (1) where a	
CC	segment of the gene has been replaced by a segment of a foreign DNA	
CC	sequence which encodes a foreign epitope or antigen. Also described are:	
CC	(1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended	
CC	assembly system of strains of Salmonella, Escherichia coli and	
CC	Enterobacteriaceae for the production of fimbriae comprising recombinant	
CC	Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)	
CC	directing recombination of a recombinant gene into the chromosome of the	
CC	homologous species; (3) directing recombination of a recombinant gene	
CC	back into the chromosome of the homologous species, replacing the native	

CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention

XX Sequence 151 AA;
SQ

Query Match 100.0%; Score 775; DB 3; Length 151;
Best Local Similarity 100.0%; Pred. No. 1.6e-68;
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLTKVAFPAIIVSGSALAGVVPQWGGGNHNGGNSGPDSTLSIYQGSANALALQ 60
Db 1 MKLTKVAFPAIIVSGSALAGVVPQWGGGNHNGGNSGPDSTLSIYQGSANALALQ 60

QY 61 SDARKSETTITQSGVNGADVGQADNSTIELTQNGFRNNAITIDQWNAKNDITVTRVYT 120
Db 61 SDARKSETTITQSGVNGADVGQADNSTIELTQNGFRNNAITIDQWNAKNDITVTRVYT 120

QY 121 HEMAHANOTASDSVWVRQVGFNNATANQY 151
Db 121 HEMAHANOTASDSVWVRQVGFNNATANQY 151

RESULT 2
AAB36347
ID AAB36347 standard; protein; 151 AA.
XX
AC AAB36347;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa::PT3#2 amino acid sequence SEQ ID NO:14.
XX
KW Salmonella; agfa; chromosomal gene replacement; fimbrial; epitope;
XX vaccine; immune response; immunogen.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW,
XX WPI: 2000-672631/65.
XX DR N-PSDB; AAC64623.
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 136; 139pp; English.
PS

XX The present invention describes a recombinant agfa gene (1) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/7AP) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention

XX Sequence 151 AA;
SQ

Query Match 92.1%; Score 714; DB 3; Length 151;
Best Local Similarity 91.1%; Pred. No. 1.8e-62;
Matches 144; Conservative 0; Mismatches 0; Indels 14; Gaps 2;

QY 1 MKLTKVAFPAIIVSGSALAGVVPQWGGGNHNGGNSGPDSTLSIYQGSANALALQ 60
Db 1 MKLTKVAFPAIIVSGSALAGVVPQWGGGNHNGGNSGPDSTLSIYQGSANALALQ 60

QY 61 SDARKSETTITQSGVNGADVGQADNSTIELTQNGFRNNAITIDQWNAKNDITVTRVYT 120
Db 61 SDARKSETTITQSGVNGADVGQADNSTIELTQNGFRNNAITIDQWNAKNDITVTRVYT 120

QY 114 LVTRVTHMAHANA-----SVWVRQVGFNNATANQY 151
Db 121 LVTRVTHMAHANA-----SVWVRQVGFNNATANQY 151

RESULT 3
AAB36355
ID AAB36355 standard; protein; 151 AA.
XX
AC AAB36355;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa::PT3#10 amino acid sequence SEQ ID NO:30.
XX
KW Salmonella; agfa; chromosomal gene replacement; fimbrial; epitope;
XX vaccine; immune response; immunogen.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX

DR N-PSDB; AAC64617.

XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.

XX Disclousure; Page 135; 139pp; English.

CC The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregate fimbriae (SEF17/TRA) nucleation depended
CC assembly system of strains of *Salmonella*, *Escherichia coli* and
CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
CC *Enterobacteriaceae* host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention

XX Sequence 151 AA;

Query Match 89.0%; Score 690; DB 3; Length 151;
Best Local Similarity 91.4%; Pred. No. 4,1e-60;
Matches 138; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

QY 1 MLLKVAFAIIVSGSALAGVVPOMGGGNNHNGGNSGPDSTLSIYOGSANAALALQ 60
DB 1 MLLKVAFAIIVSGSALAGVVPOMGGGNNHNGGNSGPDSTLSIYOGSANAALALQ 60
QY 61 SDARKSETTITGSGYGNAGADVQGADNSTIELTONGFRNNATIDQMAKXVDQVTRVVT 120
DB 61 SDARKSETTITGSGYGNAGADVQGADNSTIELTONGFRNNATIDQMAKXVDQVTRVVT 120
QY 121 HEMAHANQTAASDSSVMVRQVGFNNATANOY 151
DB 121 NNALVNOTASDSSVMVRQVGFNNATANOY 151

RESULT 6

AAW23570
ID AAW23570 standard; protein; 151 AA.

XX AAW23570;

DT 25-MAR-2003 (revised)
DT 29-SEP-1997 (first entry)

XX *Salmonella enteritidis* 27655-3b agfa.

XX Enteropathogenic bacteria; enterobacteria; *S. enteritidis*; antibody.

XX *Salmonella enteritidis*.

XX Key Location/Qualifiers

FT Misc-difference 123
XX /note= "Encoded by GCC"

PN US635617-A.

XX 03-JUN-1997.

XX 26-APR-1994; 94US-00233788.

XX 26-APR-1993; 93US-00054452.

XX (IYVT-) UNIV VICTORIA INNOVATION & DEV CORP.

XX Collinson SK, Kay WW, Doran JL;

XX WPI; 1997-309886/28.

DR N-PSDB; AAT74142.

PT Isolated *Salmonella* gene agfa - used for diagnosis of *Salmonella* or
PT enteropathogenic bacteria of the Enterobacteria family.

XX Example 2; Fig 7; 85pp; English.

CC The present sequence represents agfa encoded by the full agfa gene
CC derived from *Salmonella enteritidis* 27655-3b. The nucleic acid can be
CC used to provide diagnostic assays for *Salmonella* and/or enteropathogenic
CC bacteria of the family Enterobacteria. It can also be used to provide
CC proteins and antibodies which can be used for assays. The nucleic acid
CC hydride can be used to provide probes or primers which can specifically
CC hybridise to nucleic acid molecules from greater than 99% of *Salmonella*
CC strains that are pathogenic to warm-blooded animals relative to nucleic
CC acid molecules from virtually all other microbial organisms. (Updated on
CC 25-MAR-2003 to correct PF field.)

XX Sequence 151 AA;

Query Match 88.9%; Score 689; DB 2; Length 151;
Best Local Similarity 91.4%; Pred. No. 5.2e-60;
Matches 138; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

QY 1 MLLKVAFAIIVSGSALAGVVPOMGGGNNHNGGNSGPDSTLSIYOGSANAALALQ 60
DB 1 MLLKVAFAIIVSGSALAGVVPOMGGGNNHNGGNSGPDSTLSIYOGSANAALALQ 60
QY 61 SDARKSETTITGSGYGNAGADVQGADNSTIELTONGFRNNATIDQMAKXVDQVTRVVT 120
DB 61 SDARKSETTITGSGYGNAGADVQGADNSTIELTONGFRNNATIDQMAKXVDQVTRVVT 120
QY 121 HEMAHANQTAASDSSVMVRQVGFNNATANOY 151
DB 121 NNALVNOTASDSSVMVRQVGFNNATANOY 151

RESULT 7

AA36346
ID AAB36346 standard; protein; 151 AA.

XX AAB36346;

DT 26-FEB-2001 (first entry)

XX Agfa::PR3#1 amino acid sequence SEQ ID NO:12.

XX *Salmonella*; agfa; chromosomal gene replacement; fimbrial; epitope;
KW vaccine; immune response; immunogen.

XX *Salmonella enteritidis*.

XX *Escherichia coli*.

XX Synthetic.

XX WO200060102-A2.

XX 12-OCT-2000.
XX 05-APR-2000; 2000WO-CA000356.

PR 05-APR-1999; 99US-0127888P.
 XX (UYVI-) UNIV VICTORIA.
 PA White AP, Doran JL, Collison SK, Kay MW;
 XX MPI; 2000-672631/65.
 XX N-PSDB; AAC64622.
 DR
 XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 135; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/7AF) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;
 Query Match 84.5%; Score 655; DB 3; Length 151;
 Best Local Similarity 79.8%; Pred. No. 1.2e-56;
 Matches 134; Conservative 0; Mismatches 0; Indels 34; Gaps 2;
 QY 1 MKLLKVAAPFAIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYOGSANAALAIQ 60
 DB 1 MKLLKVAAPFAIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYOGSANAALAIQ 60
 QY 61 SDARKSETTITOSGNGADVGAGDNSTLTETLTONFRNNAITTDONAKNSDITVQYGG 120
 DB 61 SDARKSETTITOSGNGADVGAGDNSTLTETLTONFRNNAITTDONAKNSDITVQYGG 120
 QY 110 -----NYDQLVTVVTHEMAHANQTASDSVWVRQVGFGNNAATANY 151
 DB 121 NNAALVNVQDLVTRVYTHEMAH-----NNAITANY 151
 RESULT 8
 AAB36353
 ID AAB36353 standard; protein; 151 AA.
 XX AAB36353;
 AC 26-FEB-2001 (first entry)
 DT
 XX
 XX Agfa::PTJ#8 amino acid sequence SEQ ID NO:26.
 DE
 XX
 XX *Salmonella*; agfa; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.

XX
 OS *Salmonella enteritidis*.
 OS *Escherichia coli*.
 OS Synthetic.
 XX
 PN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 XX 05-APR-2000; 2000MO-CA000356.
 PF
 PR 05-APR-1999; 99US-0127888P.
 XX
 XX (UYVI-) UNIV VICTORIA.
 PA White AP, Doran JL, Collison SK, Kay MW;
 PI MPI; 2000-672631/65.
 DR N-PSDB; AAC64629.
 DR
 XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 138; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/7AF) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;
 Query Match 78.6%; Score 609; DB 3; Length 151;
 Best Local Similarity 82.1%; Pred. No. 4.2e-52;
 Matches 124; Conservative 4; Mismatches 23; Indels 0; Gaps 0;
 QY 1 MKLLKVAAPFAIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYOGSANAALAIQ 60
 DB 1 MKLLKVAAPFAIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYOGSANAALAIQ 60
 QY 61 SDARKSETTITOSGNGADVGAGDNSTLTETLTONFRNNAITTDONAKNSDITVQYGG 120
 DB 61 LVTRVYTHEMAHAGVNGADVGAGDNSTLTETLTONFRNNAITTDONAKNSDITVQYGG 120
 QY 121 HEMAHANQTASDSVWVRQVGFGNNAATANY 151
 DB 121 NNAALVNVQDLVTRVYTHEMAH-----NNAITANY 151
 RESULT 9

ID	AAB36351
XX	AAB36351 standard; protein; 151 AA.
AC	AAB36351;
XX	
DT	26-FEB-2001 (first entry)
XX	
DE	Agfa::PT#6 amino acid sequence SEQ ID NO:22.
XX	
KM	Salmonella; agfA; chromosomal gene replacement; fimbria; epitope;
KW	vaccine; immune response; immunogen.
XX	
OS	Salmonella enteritidis.
OS	Escherichia coli.
OS	Synthetic.
PN	WO200060102-A2.
XX	
PD	12-OCT-2000.
XX	
PF	05-APR-2000; 2000MO-CA000356.
XX	
PR	05-APR-1999; 99US-0127888P.
XX	
PA	(UYVI-) UNITV VICTORIA.
PI	White AP, Doran JL, Collison SK, Kay WW;
DR	WPI: 2000-672631/65.
XX	N-PESDB; AAC64627.
PT	Recombinant agfa gene having a segment replaced by a foreign DNA sequence which encodes a foreign epitope or antigen, expresses recombinant Agfa
PT	protein useful for eliciting immune response in animal.
PS	Disclosure; Page 137; 139pp; English.
XX	
CC	The present invention describes a recombinant agfa gene (1) where a
CC	sequence of the gene has been replaced by a segment of a foreign DNA
CC	sequence which encodes a foreign epitope or antigen. Also described are:
CC	(1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC	assembly system of strains of Salmonella, Escherichia coli and
CC	Enterobacteriaceae for the production of fimbriae comprising recombinant
CC	Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
CC	directing recombination of a recombinant gene into the chromosome of the
CC	homologous species; (3) directing recombination of a recombinant gene
CC	back into the chromosome of the homologous species, replacing the native
CC	copy of that gene; and (4) eliciting an immune response in an animal,
CC	comprising separating an amino acid polymer comprising a recombinant Agfa
CC	protein containing a replacement segment or segments of foreign amino
CC	acid sequence or sequences grown on a Salmonella, E. coli or
CC	Enterobacteriaceae host cell, from the host cell and introducing the
CC	polymer into the animal in conjunction with a carrier or diluent. (1) is
CC	useful for the expression of recombinant Agfa protein which is useful for
CC	eliciting an immune response in an animal. In a fimbrial presentation
CC	system the heterologous antigens are presented in high numbers (up to
CC	500,000 copies/cell), the hybrid fimbria protein possesses both the
CC	immunogenicity and adhesion properties relevant for an efficient live
CC	vaccine, the carrier fimbrial subunit proteins are usually strong
CC	immunogens, which may be important for directing an immune response
CC	against the inserted epitope, and hybrid fimbriae are easy and
CC	expensive to purify in large amount. The present sequence is given in
CC	the exemplification of the present invention
XX	
QY	Sequence 151 AA;
XX	
Query Match	78.6%; Score 609; DB 3; Length 151;
Best Local Similarity	73.6%; Pred. No. 4,2e-52;
Matches 128; Conservative	0; Mismatches 0; Indels 46; Gaps 2
db	1 MKLTKVAFAAIIVSGSALAGVPPWGCGGNNHGGNSGDPSTLSIYOYSANNAALAQ 60 1 MKLTKVAFAAIIVSGSALAGVPPWGCGGNNHGGNSGDPSTLSIYOYSANNAALAQ 60

QY 6 SDAKSEETITTOGCGNGADVQGAGDNRSTIELTLQNGFNNAATIDQMNAKYDOLVTRVYT 120
Db SDARKSETITITIGSGNGADVQAGD-----NIDQLVTRVYT 97

QY 121 HEMAH-----NOTASDSVMVRQVGFNNATANQY 151
Db 98 HEMAHADQMNAKNSDITVGQYGNNAAVLVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 10
AAB36349 ID AAB36349 standard; protein; 151 AA.
AC AAB36349;
XX
DT 26-FEB-2001 (first entry)
XX
DE AGFA::PT3#4 amino acid sequence SEQ ID NO:18.
XX
KM Salmonella; agfA; chromosomal gene replacement; fimbryn; epitope;
KM vaccine; immune response; immunogen.
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX W0200060102-A2.
EN
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000MO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX NPI; 2000-672631/65.
DR N-PSDB; AAC64625.
XX
PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 136; 139pp; English.

The present invention describes a recombinant agfa gene (I) where a segment of the gene has been replaced by a segment of a foreign DNA sequence which encodes a foreign epitope or antigen. Also described are:
(1) use of thin aggregative fimbriae (SEF17/TRF) nucleation depended assembly system of strains of Salmonella, Escherichia coli and Enterobacteriaceae for the production of fimbriae comprising recombinant Agfa, CsgA and Agfa-homologue fimbryn subunits, respectively; (2) directing recombination of a recombinant gene into the chromosome of the homologous species; (3) directing recombination of a recombinant gene back into the chromosome of the homologous species, replacing the native copy of that gene; and (4) eliciting an immune response in an animal, comprising separating an amino acid polymer comprising a recombinant Agfa protein containing a replacement segment or segments of foreign amino acid sequence or sequences grown on a Salmonella, E. coli or Enterobacteriaceae host cell, from the host cell and introducing the polymer into the animal in conjunction with a carrier or diluent. (1) is useful for the expression of recombinant Agfa protein which is useful for eliciting an immune response in an animal. In a fimbrial presentation system the heterologous antigens are presented in high numbers (up to 500,000 copies/cell), the hybrid fimbryn protein possesses both the immunogenicity and adhesion properties relevant for an efficient live vaccine, the carrier fimbrial subunit proteins are usually strong immunogens, which may be important for directing an immune response against the inserted epitope, and hybrid fimbriae are easy and inexpensive to purify in large amount. The present sequence is given in

CC the exemplification of the present invention
XX Sequence 151 AA;
SQ

Query Match 78.5%; Score 608; DB 3; Length 151;
Best Local Similarity 82.1%; Pred. No. 5.2e-52; Indels 0; Gaps 0;
Matches 124; Conservative 5; Mismatches 22; Indels 0; Gaps 0;

QY 1 MKLLKVAAPFAIVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYQGSANALALQ 60
DB 1 MKLLKVAAPFAIVSGSALAGVVPQWGGGNNHGGNSGPDVLDQVTRVTHEMAHALQ 60
QY 61 SDARKSETTITGSGYNGADVGGADNSTIELTONGFRNNATIDQNAKXYDQVTRVVT 120
DB 61 SDARKSETTITGSGYNGADVGGADNSTIELTONGFRNNATIDQNAKNSDITVGQYGG 120
QY 121 HEMAHANOTASDSSVWVROVFGNNATANQY 151
DB 121 NNALVNOTASDSSVWVROVFGNNATANQY 151

RESULT 11

AAB36350
ID AAB36350 standard; protein; 151 AA.
XX
AC AAB36350;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa::PTJ#5 amino acid sequence SEQ ID NO:20.
XX
KM Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
KW vaccine; immune response; immunogen.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UUYI-) UNIT VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WM;
PI MPI; 2000-672631/65.
DR N-PSDB; AAC64630.
XX

PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 137; 139pp; English.

XX The present invention describes a recombinant agfa gene (1) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, Cgga and Agfa-homologue fimbria subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or

CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention

SQ Sequence 151 AA;

Query Match 78.2%; Score 606; DB 3; Length 151;
Best Local Similarity 82.1%; Pred. No. 8.2e-52; Indels 0; Gaps 0;
Matches 124; Conservative 5; Mismatches 22; Indels 0; Gaps 0;

QY 1 MKLLKVAAPFAIVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYQGSANALALQ 60
DB 1 MKLLKVAAPFAIVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYQGSANALALQ 60
QY 61 SDARKSETTITGSGYNGADVGGADNSTIELTONGFRNNATIDQNAKXYDQVTRVVT 120
DB 61 SDARKYDQVTRVTHEMAHAGGADNSTIELTONGFRNNATIDQNAKNSDITVGQYGG 120
QY 121 HEMAHANOTASDSSVWVROVFGNNATANQY 151
DB 121 NNALVNOTASDSSVWVROVFGNNATANQY 151

RESULT 12

AAB36354
ID AAB36354 standard; protein; 151 AA.
XX
AC AAB36354;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa::PTJ#9 amino acid sequence SEQ ID NO:28.
XX
KM Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
KW vaccine; immune response; immunogen.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UUYI-) UNIT VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WM;
PI MPI; 2000-672631/65.
DR N-PSDB; AAC64630.
XX

PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 138; 139pp; English.

XX The present invention describes a recombinant agfa gene (1) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:

CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

XX Sequence 151 AA;

Query Match 77.4%; Score 600; DB 3; Length 151;
 Best Local Similarity 81.5%; Pred. No. 3.2e-51;
 Matches 123; Conservative 3; Mismatches 25; Indels 0; Gaps 0;

QY 1 MKLLKVAFAIIVSGSALAGVVPQWGGGNNHNGGNSGDPSTLSIYQGSANALALQ 60
 DB 1 MKLLKVAFAIIVSGSALAGVVPQWGGGNNHNGGNSGDPSTLSIYQGSANALALQ 60
 QY 61 SDARKSETTITQSGNGADVGQGDNSTIELTONGFRNNATTIDQWNAKNSDITVGYGG 120
 DB 61 SDARKSETTITQSGNGADVDQLVTRVTHEMAHAFRNATTIDQWNAKNSDITVGYGG 120
 QY 121 HEMAHANQTASDSVVRQVGFNNATANQY 151
 DB 121 NNAALVNGTASDSVVRQVGFNNATANQY 151

RESULT 13

ID AAB36348 standard; protein; 151 AA.

XX AAB36348;

XX 26-FEB-2001 (first entry)

DE Agfa::PT3#3 amino acid sequence SEQ ID NO:16.

XX *Salmonella*; agfa; chromosomal gene replacement; fimbrial; epitope;

KW vaccine; immune response; immunogen.

XX *Salmonella enteritidis*.

OS *Escherichia coli*.

OS Synthetic.

XX WO200060102-A2.

XX 12-OCT-2000.

PF 05-APR-2000; 2000WO-CA000356.

PR 05-APR-1999; 99US-0127888P.

XX (UYVI-) UNIV VICTORIA.

XX White AP, Doran JL, Collison SK, Kay WW,

XX WPI; 2000-672631/65.

DR N-PSDB; AAC64624.

XX

PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.

PS Disclosure; Page 136; 139p; English.

CC The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

XX Sequence 151 AA;

Query Match 74.2%; Score 575; DB 3; Length 151;
 Best Local Similarity 81.5%; Pred. No. 9.5e-49;
 Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAFAIIVSGSALAGVVPQWGGGNNHNGGNSGDPSTLSIYQGSANALALQ 60
 DB 1 MKLLKVAFAIIVSGSALAGVVPQWGGGNNHNGGNSGDPSTLSIYQGSANALALQ 60
 QY 61 SDARKSETTITQSGNGADVGQGDNSTIELTONGFRNNATTIDQWNAKNSDITVGYGG 120
 DB 61 SDARKSETTITQSGNGADVGQGDNSTIELTONGFRNNATTIDQWNAKNSDITVGYGG 120
 QY 121 HEMAHANQTASDSVVRQVGFNNATANQY 151
 DB 121 NNAALVNGTASDSVVRQVGFNNATANQY 151

RESULT 14

ID AAB36343 standard; protein; 151 AA.

XX AAB36343;

XX 26-FEB-2001 (first entry)

DE *Escherichia coli* CsgA amino acid sequence SEQ ID NO:7.

XX *Salmonella*; agfa; chromosomal gene replacement; fimbrial; epitope;

KW vaccine; immune response; immunogen.

XX *Escherichia coli*.

XX WO200060102-A2.

XX 12-OCT-2000.

XX

PF 05-APR-2000; 2000MO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 XX (UYVI-) UNIV VICTORIA.
 PA
 XX White AP, Doran JL, Collison SK, Kay WM;
 PI
 XX MPI: 2000-672631/65.
 DR N-PSDB; AAC64619.
 XX
 XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 135; 139pp; English.

XX The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/7AP) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2) the
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal, a
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

XX Sequence 151 AA;

XX Query Match 68.8%; Score 533; DB 3; Length 151;

XX Best Local Similarity 70.2%; Pred. No. 1.3e-44;

XX Matches 106; Conservative 18; Mismatches 27; Indels 0; Gaps 0;

QY 1 MLLKVAAPFAIVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYOGSANAALALQ 60
 DB 1 MLLKVAAPFAIVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYOGGNSALALQ 60
 QY 61 SPARKSETTITGSGYNGADVGQADNSTIETLTONGFRRNATIDQNNAKNYDLTVRYVT 120
 DB 61 TDARNSDLTITGHHGGNGADVGQSDSDSLDTRGFNGSATIDQNNKNSERTVYQFGG 120
 QY 121 HEMAHANOTASDSVYVRYGVFGNNATANY 151
 DB 121 GNGAAVDOTASNSVYVTVGVFGNNATANY 151

XX RESULT 15

XX ABR82651

XX ABR82651 standard; protein; 151 AA.

XX ABR82651;

XX 04-DEC-2003 (first entry)

XX E. coli CsgA subunit 15 kDa protein.

KW Plasma protein; immune response; antibacterial; vaccine; gene therapy.

XX *Escherichia coli*.

XX WO2003064446-A2.

XX 07-AUG-2003.

XX 30-JAN-2003; 2003WO-EP000943.

XX 31-JAN-2002; 2002GB-00002275.

XX (HANS-) HANSA MEDICAL RES AB.

XX Bjoerck L, Olsen A, Wikeroem M, Herwald H;

XX MPI: 2003-646136/61.

XX N-PSDB; ACP36153.

XX New isolated peptide capable of binding a mammalian plasma protein,
 PT useful in the manufacture of a medicament for the prevention and/or
 PT treatment of a bacterial infection, such as *Escherichia coli*, *Salmonella*
 PT or *Shigella* infections.

XX Disclosure; Page 41-42; 42pp; English.

XX The invention relates to an isolated peptide capable of binding a
 CC mammalian plasma protein or of generating an immune response in a mammal
 CC selected from sequences shown in ABR82642, ABR82648-49. The peptide or
 CC antibody is useful for treating a bacterial infection in a human or
 CC animal or in the manufacture of a medicament for the prophylactic
 CC treatment of a bacterial infection, such as *Escherichia coli*, *Salmonella*
 CC or *Shigella* infection. The peptide that is immobilized on a solid support
 CC is also useful as a reagent for determining the ability of a plasma
 CC protein to bind to bacteria. The present sequence represents an *E. coli*
 CC 15 kDa protein

XX Sequence 151 AA;

XX Query Match 69.1%; Score 528; DB 7; Length 151;

XX Best Local Similarity 69.5%; Pred. No. 4.2e-44;

XX Matches 105; Conservative 18; Mismatches 26; Indels 0; Gaps 0;

QY 1 MLLKVAAPFAIVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYOGSANAALALQ 60
 DB 1 MLLKVAAPFAIVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYOGGNSALALQ 60
 QY 61 SPARKSETTITGSGYNGADVGQADNSTIETLTONGFRRNATIDQNNAKNYDLTVRYVT 120
 DB 61 TDARNSDLTITGHHGGNGADVGQSDSDSLDTRGFNGSATIDQNNKNSERTVYQFGG 120
 QY 121 HEMAHANOTASDSVYVRYGVFGNNATANY 151
 DB 121 GNGAAVDOTASNSVYVTVGVFGNNATANY 151

XX Search completed: March 11, 2004, 18:33:44

XX Job time : 45.9 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:25:54 ; Search time 12.4 Seconds

(without alignments)
628.671 Million cell updates/sec

Title: US-09-543-407-24

Sequence: 1 MLKLKVAAPAAIVSGSALA.....DSSVMRYQFGNNATFANY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Issued Patents AA:*

- 1: /cgn2_6/prodata/2/iaa/5A_COMB.pep:*
- 2: /cgn2_6/prodata/2/iaa/5B_COMB.pep:*
- 3: /cgn2_6/prodata/2/iaa/6A_COMB.pep:*
- 4: /cgn2_6/prodata/2/iaa/6B_COMB.pep:*
- 5: /cgn2_6/prodata/2/iaa/CCTUS_COMB.pep:*
- 6: /cgn2_6/prodata/2/iaa/backfillseq1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	689	88.9	151	1	US-08-233-788A-59
2	505	65.2	120	1	US-08-233-788A-57
3	90.5	11.7	673	3	US-09-196-387-8
4	90.5	11.7	673	4	US-09-841-835-8
5	90.5	11.7	949	3	US-09-196-387-10
6	90.5	11.7	949	4	US-09-841-835-10
7	90.5	11.7	1327	3	US-09-196-387-2
8	90.5	11.7	1327	4	US-09-841-835-2
9	90.5	11.7	1327	4	US-09-972-115A-8
10	89.5	11.5	738	3	US-08-864-038A-3
11	87	11.2	1156	4	US-09-198-452A-171
12	84.5	10.9	339	4	US-09-252-991A-32096
13	83	10.7	889	4	US-09-336-447A-15
14	81.5	10.5	720	1	US-07-731-157A-2
15	81.5	10.5	720	2	US-08-541-780-2
16	80.5	10.4	389	4	US-09-134-001C-4316
17	80	10.3	2736	4	US-09-252-991A-30227
18	79.5	10.3	881	4	US-09-489-039A-12003
19	79	10.2	1415	4	US-09-252-991A-26438
20	79	10.2	2123	3	US-08-568-685A-10
21	78.5	10.1	589	4	US-09-489-039A-7849
22	78.5	10.1	745	4	US-09-336-115C-6
23	77	9.9	415	4	US-09-025-769B-280
24	77	9.9	444	1	US-07-881-075-3
25	77	9.9	444	1	US-08-120-827-3
26	77	9.9	444	1	US-08-478-675-3
27	77	9.9	1690	4	US-09-595-684B-39

28	76.5	9.9	227	4	US-09-328-352-7735	Sequence 7735, Ap
29	76	9.8	2680	4	US-09-489-039A-7973	Sequence 7973, Ap
30	75.5	9.7	238	4	US-09-495-880A-42	Sequence 42, Appl
31	75.5	9.7	645	4	US-09-919-172-41	Sequence 41, Appl
32	75.5	9.7	789	3	US-08-960-780-6	Sequence 6, Appl
33	75.5	9.7	789	3	US-09-073-898-6	Sequence 6, Appl
34	75.5	9.7	789	4	US-08-850-351A-6	Sequence 6, Appl
35	75.5	9.7	1036	4	US-09-206-942-73	Sequence 73, Appl
36	75.5	9.7	1051	4	US-09-134-001C-5005	Sequence 5005, Ap
37	75.5	9.7	1477	1	US-08-038-682-4	Sequence 4, Appl
38	75.5	9.7	1477	2	US-08-302-832-4	Sequence 4, Appl
39	75.5	9.7	1477	2	US-08-530-198-4	Sequence 4, Appl
40	75.5	9.7	1477	2	US-08-469-880-4	Sequence 4, Appl
41	75.5	9.7	1477	2	US-08-476-4	Sequence 4, Appl
42	75.5	9.7	1477	2	US-08-617-697-4	Sequence 4, Appl
43	75.5	9.7	1477	3	US-08-719-641-4	Sequence 4, Appl
44	75.5	9.7	1477	3	US-09-206-942-71	Sequence 71, Appl
45	75	9.7	211	1	US-08-276-852-34	Sequence 34, Appl

ALIGNMENTS

RESULT 1
US-08-233-788A-59
Sequence 59, Application US/08233788A
Patent No. 5635617
GENERAL INFORMATION:
APPLICANT: Doran, James L.
APPLICANT: Kay, William W.
APPLICANT: Colinsion, Karen S.
APPLICANT: Clouthier, Sharon C.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
OF SALMONELLA
NUMBER OF SEQUENCES: 61
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed and Berry
STREET: 6300 Columbia Center, 701 Fifth Avenue
City: Seattle
STATE: Washington
COUNTRY: U.S.A.
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233,788A
FILING DATE: 26-APR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: King, Joshua
REGISTRATION NUMBER: 35,570
REFERENCE/DOCKET NUMBER: 920043.403C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEDANBERY
INFORMATION FOR SEQ ID NO: 59:
SEQUENCE CHARACTERISTICS:
LENGTH: 151 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-233-788A-59

Query Match 88.9%; Score 689; DB 1; Length 151;
Best Local Similarity 91.4%; Pred. No. 1.8e-60;
Matches 138; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

QY 1 MLKLKVAAPAAIVSGSALAGVPPWGGGNNNGGNSGPPSTISITVYGSANNAALAO 60
|||||

Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNNHGGSSGPDSTLSTIYGSANAAALQ 60
QY 61 SARKSETTTTGGYNGADVGAGANSTTELTONGFRNNATTIDOMAKNYDQVTRVVT 120
Db 61 SDRKSETTTTGGYNGADVGAGANSTTELTONGFRNNATTIDOMAKNSDITVGYG 120
QY 121 HEMAHANQTASDSVAVRQVGFNNATANOY 151
Db 121 NNPAIVNQTSADSSVAVRQVGFNNATANOY 151

RESULT 2

US-08-233-788A-57
Sequence 57, Application US/08233788A
Patent No. 5635617

GENERAL INFORMATION:

APPLICANT: Doran, James L.
APPLICANT: Kay, William W.
APPLICANT: Collinson, Karen S.
APPLICANT: Clouthier, Sharon C.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
TITLE OF INVENTION: OF SALMONELLA
NUMBER OF SEQUENCES: 61
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed and Berry
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: U.S.A.
ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233, 788A
FILING DATE: 26-APR-1994
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: King, Joshua
REGISTRATION NUMBER: 35,570
REFERENCE/DOCKET NUMBER: 920043.403C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEDDANBERRY
INFORMATION FOR SEQ ID NO: 57:

SEQUENCE CHARACTERISTICS:

LENGTH: 120 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-233-788A-57

Query Match

Best Local Similarity 65.2%; Score 505; DB 1; Length 120;
Pred. No. 1,7e-42;
Matches 99; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

QY 22 VVPWGGGNNHGGSSGPDSTLSTIYGSANAAALQSDARKSETTTTGGYNGADV 81
Db 1 VVPWGGGNNHGGSSGPDSTLSTIYGSANAAALQSDARKSETTTTGGYNGADV 81
QY 82 GGGADNSTTELTONGFRNNATTIDOMAKNYDQVTRVVTHEMAHANOQTASDS 133
Db 61 GGGADNSTTELTONGFRNNATTIDOMAKNSDITVGYGNNALVNOQTASDS 112

RESULT 3

US-09-196-387-8
Sequence 8, Application US/09196387
Patent No. 6277613

GENERAL INFORMATION:

APPLICANT: de Lange, Titia

APPLICANT: Smith, Susan

TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS

TITLE OF INVENTION: OF USE THEREOF

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson

STREET: 411 Hackensack Avenue, 4th Floor

CITY: Hackensack

STATE: New Jersey

COUNTRY: USA

ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/196,387

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/095,225

FILING DATE: June 10, 1998

ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.

REGISTRATION NUMBER: 26,742

REFERENCE/DOCKET NUMBER: 600-1-230 CIP1

TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800

TELEFAX: 201-343-1684

TELEX: 133521

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:
LENGTH: 673 amino acids

TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

US-09-196-387-8

Query Match

Best Local Similarity 11.7%; Score 90.5; DB 3; Length 673;
Pred. No. 1.1;
Matches 42; Conservative 16; Mismatches 61; Indels 29; Gaps 6;

QY 6 VAAFAAI-VVSGSALAGVVPQWGGGNNHGGSSGPDSTLSTIYGSANAAALQSDAR 64
Db 99 VAAAPVPAVSTSSAAGVAPNPAGSGNNSSSSPTSS-SSSSPSSPGSLAESPEAA 157

QY 65 KSETTTT---OSGYNGADVGAGANSTTELTONGFRNNATTIDOMAKNYDQVTRVVT 120
Db 158 GVSSTAFLGPGAAAGGTGVPVAVSGALRELLACRG-----DVSRYRLVDA--- 204

QY 121 HEMAHANQTASD-----SSVAVRQVGF 143
Db 205 -----ANVAKDMAGRKSSPLHFAFG 227

US-09-841-835-8

RESULT 4

Sequence 8, Application US/09841835
Patent No. 6506587

GENERAL INFORMATION:

APPLICANT: de Lange, Titia

APPLICANT: Smith, Susan

TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS

TITLE OF INVENTION: OF USE THEREOF

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson

STREET: 411 Hackensack Avenue, 4th Floor

CITY: Hackensack

STATE: New Jersey

COUNTRY: USA


```

      ZIP: 07601
      COMPUTER READABLE FORM:
      MEDIUM TYPE: Floppy disk
      COMPUTER: IBM PC compatible
      OPERATING SYSTEM: PC-DOS/MS-DOS
      SOFTWARE: Patentin Release #1.0, Version #1.30
      CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/09/841,835
      FILING DATE:
      CLASSIFICATION:
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: 09/196,387
      FILING DATE:
      ATTORNEY/AGENT INFORMATION:
      NAME: Jackson Esq., David A.
      REGISTRATION NUMBER: 26,742
      REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
      TELECOMMUNICATION INFORMATION:
      TELEPHONE: 201-487-5800
      TELEFAX: 201-343-1684
      TELEX: 133521
      INFORMATION FOR SEQ ID NO: 8:
      SEQUENCE CHARACTERISTICS:
      LENGTH: 673 amino acids
      TYPE: amino acid
      TOPOLOGY: linear
      MOLECULE TYPE: protein
      US-09-841-835-8

Query Match          11.7%; Score 90.5; DB 4; Length 673;
Best Local Similarity 28.4%; Pred.No.1.1;
Matches 42; Conservative 16; Mismatches 61; Indels 29; Gaps 6

QY   6 VAAFAAI-VSSGSALAGVPPWGGGNGNHNCGNSGSPDSTLSIYGYSGANNAALALQSDAR 64
    |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB   99 VAAAVVAVRVTSSSAQAAPNPAGSGSNSSPSSSPTSS-SSSPS$FGSSLAEPEAA 157
    QY   65 KSETIT---QSGYGNGADVQGADNSTIELTONGFRNNATIDDMANNYDQLTVRYVT 120
        ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB   158 GVSSAPAPGPAGAAGGTGVPAVSAGLRLELACRMG-----DVRVRKLVDYDA---- 204
        :|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY   121 HEMAHANTASD-----SSVMRYQGFPG 143 .
        |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB   205 -----ANVNAKDMGRKSSPLHFAAGF 227
        ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

RESULT 5
US-09-196-387-10
Sequence 10, Application US/09196387
Patent No. 6277613
GENERAL INFORMATION:
APPLICANT: de Lange, Titia
APPLICANT: Smith, Susan
TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRP1 AND METHODS
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue, 4th Floor
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/196,387
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:

```

```

APPLICATION NUMBER: 09/095,225
FILING DATE: June 10, 1998
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 949 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-196-387-10

Query Match      11.7%; Score 90.5; DB 3; Length 949;
Best Local Similarity 28.4%; Pred. No. 1.8;
Matches 42; Conservative 16; Mismatches 61; Indels 29; Gaps 6

Qy      VAAFAAI-VVSGSLAGVPQMGGGNGNNGSSGPSTLSIVYQSANAALQSPAR 64
        |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      99 VAAPVPVPAVTSSTASAGAAPNPASSGNSSSSSSPSS-SSSSPSSPGSILASPRA 157
        ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

Qy      65 KSEFTIT---QSGYGNCADVGGAGADNSTIELTONGFNNAITIDQMAKNYDOLTVRYVT 120
        :|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      158 GVSSTAPLPGGAAPGTGCVPVAVSGLARELLERCRNG-----DVSRVRLVDVA--- 204
        :|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

Qy      121 HEMAHANQTASD-----SSWVRVOYRG 143
        |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      205 -----ANVNAKDMAGKRKSSPLHFAAFG 227
        ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

RESULT 6
US-09-841-835-10
Sequence 10, Application US/09841835
Patent No. 6506587
GENERAL INFORMATION:
APPLICANT: de Lange, Titia
APPLICANT: Smith, Susan
TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
TITLE OF INVENTION: OF USE THEREOF
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue, 4th floor
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/841,835
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/196,387
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 10:

```


QY 121 HEMAHANQTSAD-----SSVMROVFG 143
 Db 205 -----ANVNAKDMAGRKSSPLHFAAGFG 227

RESULT 9

US-09-972-115A-8
 ; Sequence 8, Application US/09972115A
 ; Patent No. 6599728
 ; GENERAL INFORMATION:
 ; APPLICANT: Geron Corporation
 ; APPLICANT: Gregg, Morin B.
 ; APPLICANT: Walter, Funk D.
 ; APPLICANT: Mieczyslaw, Piatyzek A.
 ; TITLE OF INVENTION: A Second Mammalian Telomerase
 ; FILE REFERENCE: 080/003C
 ; CURRENT APPLICATION NUMBER: US/09/972,115A
 ; PRIOR FILING DATE: 2001-10-05
 ; PRIOR APPLICATION NUMBER: US 60/128,577
 ; PRIOR FILING DATE: 2000-04-10
 ; PRIOR APPLICATION NUMBER: US 60/129,123
 ; PRIOR FILING DATE: 1999-04-13
 ; NUMBER OF SEQ ID NOS: 64
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 8
 ; LENGTH: 1327
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-972-115A-8

Query Match 11.7%; Score 90.5; DB 4; Length 1327;
 Best Local Similarity 28.4%; Pred. No. 2.8;
 Matches 42; Conservative 16; Mismatches 61; Indels 29; Gaps 6;

QY 6 VAAFAAI-VVSGALGVVPQMGCGGNHNGGSSGPDSTLSIYQGSANALALQSDAR 64
 Db 99 VAAAPVPAVSTASAGVAPNPAAGSNNSSSSSTSS-SSSSPSSPSSSLAESPEEA 157
 QY 65 KSEETIT-----QSGYGAGADVGGADNRTLETQNGFRNATIDOWNAKNYDQLYTRVVT 120
 Db 158 GVSSTAPLPGGAAGPRTGVPAVSGALRELLIACRNG-----DVSRYKRLVDA--- 204
 QY 121 HEMAHANQTSAD-----SSVMROVFG 143
 Db 205 -----ANVNAKDMAGRKSSPLHFAAGFG 227

RESULT 10

US-08-864-038A-3
 ; Sequence 3, Application US/08864038A
 ; Patent No. 6001592
 ; GENERAL INFORMATION:
 ; APPLICANT: Kunio NAKASHIMA et al.
 ; TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE CDNA, VECTOR
 ; TITLE OF INVENTION: CONTAINING SAID CDNA, HOST CELLS TRANSFORMED WITH SAID
 ; TITLE OF INVENTION: VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING
 ; TITLE OF INVENTION: SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBODY
 ; TITLE OF INVENTION: TO SAID POLYPEPTIDE
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: 812-5 Hirano
 ; STREET: Ieshinden
 ; CITY: Tsuru-city
 ; STATE: Mie-prefecture
 ; COUNTRY: JAPAN
 ; ZIP: 514-01
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: Microsoft Windows 95
 ; SOFTWARE: Word Perfect 6.1
 ; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/864,038A
 FILING DATE: May 28, 1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: JP 8-184459
 FILING DATE: 15-July-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: C. Bruce Hamburg
 REGISTRATION NUMBER: 22,389
 REFERENCE/DOCKET NUMBER: F-5610
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212)986-2340
 TELEFAX: (212)953-7733
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 738
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 ORIGINAL SOURCE:
 ORGANISM: Pinctada fucata
 CELL TYPE: mantle epithelial cell
 FEATURE:
 NAME/KEY: peptide
 LOCATION: from 1 to 738
 IDENTIFICATION METHOD: E (by experiment)

US-08-864-038A-3

Query Match 11.5%; Score 89.5; DB 3; Length 738;
 Best Local Similarity 35.4%; Pred. No. 1.6;
 Matches 29; Conservative 4; Mismatches 26; Indels 23; Gaps 3;

QY 3 LKVAFAAI-VVSGALGVVPQMGCGGNHNGGSSGPDSTLSIYQGSANALALQSD 62
 Db 419 LKSSASASASASASASAC-----GGGGGANGGNGCGG-----GGAGALAL----- 460
 QY 63 ARKSETTITQSGYGAGADVGG 84
 Db 461 -----ALALAAAGAGGGLGGGG 477

RESULT 11

US-09-198-452A-171
 ; Sequence 171, Application US/09198452A
 ; Patent No. 6559294
 ; GENERAL INFORMATION:
 ; APPLICANT: Griffiths, R.
 ; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragment
 ; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prev
 ; FILE REFERENCE: 9710-003-999
 ; CURRENT APPLICATION NUMBER: US/09/198,452A
 ; CURRENT FILING DATE: 1998-11-24
 ; NUMBER OF SEQ ID NOS: 6849
 ; SEQ ID NO 171
 ; LENGTH: 1156
 ; TYPE: PRT
 ; ORGANISM: Chlamydia pneumoniae
 US-09-198-452A-171

Query Match 11.2%; Score 87; DB 4; Length 1156;
 Best Local Similarity 32.0%; Pred. No. 5.1;
 Matches 32; Conservative 14; Mismatches 34; Indels 20; Gaps 5;

QY 44 TISIYQGSANALALQSDARKEETITQSGYGAGADVGGADNRTI-----ELTQ 94
 Db 859 TVSPEDYAAVQALAL--ATVRHESLIV-STYGLAQSGQTSSKVTTLMRDLHAEVLE 915
 QY 95 NGFRNATIDOWNAKNYDQLYTRVVTHEMAHANQTSADSS 134
 Db 916 MG-----VETRLNRSDILHRV--HSVTLHSLDSDSS 947

RESULT 12

US-09-252-991A-32096
; Sequence 32096, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 32096
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-32096

Query Match 10.9%; Score 84.5; DB 4; Length 339;
Best Local Similarity 21.5%; Pred. No. 1.8; Indels 25; Gaps 5;
Matches 34; Conservative 32; Mismatches 67

QY 5 KVAAPAAIVSGSALAGVPPQMG-----GNHNGGSSGPDSTLIYQGSANAAL 57
DB 58 KVSNEFTL--NNASVSGSIKIDASGVGVNVAAGNNQQAALASADASFVETATAS- 114
QY 58 ALQSDARKSETTITQSGYGCADVCGAGDNSTIELTONGFFNATIQMNAKYDQ---- 113
DB 115 -----TSVLOSQSGNTLNNYSNPNTASLNSANNVSGNLGVNV--AAGNFOQKND 163
QY 114 LVTRVTHEMAHANTASDSSVWROYGFGNATANY 151
DB 164 LAAVNSGQYSTAGSASQSTGTGTTVNSANYATGCTY 201

RESULT 13
US-09-336-447A-15
; Sequence 15, Application US/09336447A
; Patent No. 6310190
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USPA1 AND USPA2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/09/336.447A
; CURRENT FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
; LENGTH: 889
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-336-447A-15

Query Match 10.7%; Score 83; DB 4; Length 889;
Best Local Similarity 27.9%; Pred. No. 8.9;
Matches 39; Conservative 16; Mismatches 51; Indels 34; Gaps 8;
QY 28 GGGHNGGSSGPDSTLIYQ-----YSANAALQSDARKSETTITQSGY-----GNGA 79
DB 69 GAGHNNVVGSGAHSGILGKMTVNGYTSIVGYNETQDGFVGGYKMLAKGNYT 128
QY 80 DVGG-----ADNSTIELTONGFFN-----NATID-----QMNANVQOLVTRVTHEMAH 125
DB 129 FVGGGYKMLAEGDNATT---AGFRANLAEGDNATTIAGGFENRAGIDSVVSG-----GY 179

QY 126 ANQATSDSSVWROYGFGNN 145
DB 180 ANQATGESSTV---AGGSNN 196

RESULT 14
US-07-731-157A-2
; Sequence 2, Application US/07731157A
; Patent No. 5457032
; GENERAL INFORMATION:
; APPLICANT: Quax, Wilhelmus J.
; APPLICANT: Misser, Onno
; APPLICANT: Van der Laan, Jan M.
; APPLICANT: Lenting, Herman B.M.
; TITLE OF INVENTION: Mutated beta-lactam acylase genes
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: COOLEY GODWARD CASTRO HUDDLESON & TATUM
; STREET: FIVE PALO ALTO SQUARE, 4TH FLOOR
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/731.157A
; FILING DATE: 19910509
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 90200962
; FILING DATE: 18-APR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: RAE-VENTER PH.D., BARBARA
; REGISTRATION NUMBER: 32,750
; REFERENCE/DOCKET NUMBER: GBRO-027/00US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-494-7622
; TELEFAX: 415-857-0663
; TELEX: 380816 COOLEY PA
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 720 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-731-157A-2

Query Match 10.5%; Score 81.5; DB 1; Length 720;
Best Local Similarity 25.8%; Pred. No. 9.5;
Matches 46; Conservative 17; Mismatches 56; Indels 59; Gaps 9;

QY 1 MKLLKVAAPAAIVSGSALAGV-----PQ-----WGCGGNNGGG 36
DB 2 LVTRHRAASALVWATYTGAPAVAFALAEPTSPQPIAYKRSNEILMDG---YGVV 57
QY 37 NSSGPDSTLIYQGSANA-----ALALQSDARKSETTITQSGYGCADV--GQGDNST 89
DB 58 HLYGVAPAFGFGYGAQARSQGDNITRLYGEAR-----GKGAEYWGPDYEQT 106
QY 90 IELTONGFFNATIIDQNAK-----NYDQVTRVTHEMAHANTASDSSVWROY 140
DB 107 VWLITGVBERA--QGYTAQSDPFRANDAPARAIGI-----NAYAQNPDDISPDVQOV 158

RESULT 15
US-08-541-780-2
; Sequence 2, Application US/08541780
; Patent No. 5935831
; GENERAL INFORMATION:

APPLICANT: Quax, Wilhelmus J.
APPLICANT: Misset, Onno
APPLICANT: Van der Laan, Jan M.
APPLICANT: Lenting, Herman B.M.
TITLE OF INVENTION: Mutated beta-lactam acylase genes
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: COOLEY GODWARD CASTRO HUDDLESON & TATUM
STREET: FIVE PALO ALTO SQUARE, 4TH FLOOR
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/541,780
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/731,157
FILING DATE:
APPLICATION NUMBER: EP 90200962
FILING DATE: 18-APR-1990
ATTORNEY/AGENT INFORMATION:
NAME: RAE-VENTER PH.D., BARBARA
REGISTRATION NUMBER: 32,750
REFERENCE/DOCKET NUMBER: GBRO-027/00US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-494-7622
TELEFAX: 415-857-0663
TELEX: 380816 COOLEY PA
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 720 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-541-780-2

Query Match 10.5%; Score 81.5; DB 2; Length 720;
Best Local Similarity 25.8%; Pred. No. 9.5;
Matches 46; Conservative 17; Mismatches 56; Indels 59; Gaps 9;
QY 1 MKLLKVAAPAAIVSGSALAGV-----PQ-----WGGGNNHNGG 36
DB 2 LRLVLRRAASALVWATVIGLAPVAFALAEPTSTPCAPIAVKPRSNELMDG---YGVV 57
QY 37 NSSGPDSTLSIYOYGSANA-----ALALQSDARKSETTITQSGYNGADV--GQGADNGT 89
DB 58 HLYGVADAPSAFYGYGWAQARSGQDNILRLYGEAR-----GKGAHYWGPDYEQTT 106
QY 90 IELTONGFRNNATTIDQNAK-----NYDQLVTRVVTHEMAHANOATASDSSVMVRQV 140
DB 107 VWLLTNGVERA--QQWVAQOSPDPRANLDAFAGI---NVAQGNPDDISPDVRQV 158

Search completed: March 11, 2004, 18:44:53
Job time : 13.4 secs

THIS PAGE BLANK (USPTO)

Db 78 GNSNSV-----GRDIGKQSGAGNSAIFQEGTSGDVELQGTSTNGAVPSGMMNTN 129
 Qy 111 YQVLTVRVTHEMALAN-----QTASDSSVWTRQVGFNNATANO 150
 Db 130 DPGVFNKITQDSSSSNGSKVSVI QDKRNNVFSIKQNTGNTSVNQ 174

RESULT 2 US-10-032-585-7876

Sequence 7876, Application US/10032585
 Publication No. US20030180953A1
 GENERAL INFORMATION:
 APPLICANT: Terry, Roemer D.
 APPLICANT: Bo, Jiansg
 APPLICANT: Charles, Boone
 APPLICANT: Howard, Bussey
 TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
 FILE REFERENCE: 10182-005-999
 CURRENT APPLICATION NUMBER: US/10/032,585
 CURRENT FILING DATE: 2001-12-20
 NUMBER OF SEQ ID NOS: 8000
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO 7876
 LENGTH: 688
 TYPE: PRT
 ORGANISM: Candida albicans
 US-10-032-585-7876

Query Match 11.7%; Score 91; DB 14; Length 688;
 Best Local Similarity 23.7%; Pred. No. 1.1;
 Matches 28; Conservative 25; Mismatches 59; Indels 6; Gaps 2;

Qy 30 GNNHGGGSSGPDSTLSIYQGSANALALQSDARKSETTITOGSGNGADVGGADNST 89
 Db 488 GNNHGGGSSGPDSTLSIYQGSANALALQSDARKSETTITOGSGNGADVGGADNST 89
 Qy 90 IELTONGFRNATIDOMNANKYDOLVTRVTHEMAHANQTPASDSSVWTRQVGFNNAT 147
 Db 545 KSKKSNFDDNNSNS---ALNNLDSKSLKINTNEITNISFTSNSSPVIMLNHGRSS 599

RESULT 3 US-09-841-835-8

Sequence 8, Application US/09841835
 Patent No. US20020076795A1
 GENERAL INFORMATION:
 APPLICANT: de Lange, Titia
 APPLICANT: Smith, Susan
 TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
 TITLE OF INVENTION: OF USE THEREOF
 NUMBER OF SEQUENCES: 12
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Klauber & Jackson
 STREET: 411 Hackensack Avenue, 4th Floor
 CITY: Hackensack
 STATE: New Jersey
 COUNTRY: USA
 ZIP: 07601
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/841,835
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 09/196,387
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Jackson Esq., David A.
 REGISTRATION NUMBER: 26,742

REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 201-487-5800
 TELEFAX: 201-343-1684
 TELEX: 133521
 INFORMATION FOR SEQ ID NO: 8:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 673 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-09-841-835-8

Query Match 11.7%; Score 90.5; DB 9; Length 673;
 Best Local Similarity 28.4%; Pred. No. 1.2;
 Matches 42; Conservative 16; Mismatches 61; Indels 29; Gaps 6;

Qy 6 VAAFAAI-VVSGSALAGVVPQWGGGNNHGGNNSGPDSTLSIYQGSANALALQSDAR 64
 Db 99 VAAAPVPAVSTSSAAGVAPNPAAGSGSNSSPSSPTSS-SSSPSSPGSSLAESPAA 157
 Qy 65 KSETTIT---QSGYNGADVQGDNSTIELTONGFRNATIDOMNANKYDOLVTRVVT 120
 Db 158 GVSSTAPLPGGAAGPTGVPAVSGALRELLFACRNG-----DVSRRVRLVDA--- 204
 Qy 121 HEMAHANQTPASD-----SSVWTRQVGF 143
 Db 205 -----ANVANAKDMAGRKSSPLHPAAGFG 227

RESULT 4 US-09-841-835-10

Sequence 10, Application US/09841835
 Patent No. US20020076795A1
 GENERAL INFORMATION:
 APPLICANT: de Lange, Titia
 APPLICANT: Smith, Susan
 TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
 TITLE OF INVENTION: OF USE THEREOF
 NUMBER OF SEQUENCES: 12
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Klauber & Jackson
 STREET: 411 Hackensack Avenue, 4th Floor
 CITY: Hackensack
 STATE: New Jersey
 COUNTRY: USA
 ZIP: 07601
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/841,835
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 09/196,387
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Jackson Esq., David A.
 REGISTRATION NUMBER: 26,742
 REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 201-487-5800
 TELEFAX: 201-343-1684
 TELEX: 133521
 INFORMATION FOR SEQ ID NO: 10:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 949 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein

US-09-841-835-10

Query Match 11.7%; Score 90.5; DB 9; Length 949;
 Best Local Similarity 28.4%; Pred. No. 1.9;
 Matches 42; Conservative 16; Mismatches 61; Indels 29; Gaps 6;

QY 6 VAAFAAI-VVSGSALAGVVPQWGGGNNHGGNSGSPDSTLSIYQGSANALALQSDAR 64
 DB 99 VAAAPVPAVSTSSAGVAPNPAGSGSNNSSPSSSSPTSS-SSSSPSSPGSSLAESPFAA 157

QY 65 KSEITIT-----OSGVNGADVQGDNSITIELTONGFRNNATIDQNNAKNYDQLYTVRYVT 120
 DB 158 GVSSTAPLPGAGAGCTGVPVAVSGALRELBACRNG-----DVSRRVRLVDA---- 204

QY 121 HEMAHANOTASD-----SSVMVROYVFG 143
 DB 205 -----ANVNAKDMAGRKSSPLHFAAGFG 227

RESULT 5
 US-09-841-835-2
 ; Sequence 2, Application US/09841835
 ; Patent No. US20020076795A1

GENERAL INFORMATION:
 APPLICANT: de Lange, Titia
 APPLICANT: Smith, Susan
 TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
 TITLE OF INVENTION: OF USE THEREOF
 NUMBER OF SEQUENCES: 12
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Klauber & Jackson
 STREET: 411 Hackensack Avenue, 4th Floor
 CITY: Hackensack
 STATE: New Jersey
 COUNTRY: USA

ZIP: 07601
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/841,835
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 09/196,387
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Jackson Esq., David A.
 REGISTRATION NUMBER: 26,742
 REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 201-487-5800
 TELEFAX: 201-343-1684
 TELEX: 133521

INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1327 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHETICAL: NO
 US-09-841-835-2

Query Match 11.7%; Score 90.5; DB 9; Length 1327;
 Best Local Similarity 28.4%; Pred. No. 2.9;
 Matches 42; Conservative 16; Mismatches 61; Indels 29; Gaps 6;

QY 6 VAAFAAI-VVSGSALAGVVPQWGGGNNHGGNSGSPDSTLSIYQGSANALALQSDAR 64
 DB 99 VAAAPVPAVSTSSAGVAPNPAGSGSNNSSPSSSSPTSS-SSSSPSSPGSSLAESPFAA 157

QY 65 KSEITIT-----OSGVNGADVQGDNSITIELTONGFRNNATIDQNNAKNYDQLYTVRYVT 120
 DB 158 GVSSTAPLPGAGAGCTGVPVAVSGALRELBACRNG-----DVSRRVRLVDA---- 204

QY 121 HEMAHANOTASD-----SSVMVROYVFG 143
 DB 205 -----ANVNAKDMAGRKSSPLHFAAGFG 227

RESULT 6
 US-09-972-115A-8
 ; Sequence 8, Application US/0972115A
 ; Publication No. US20030032769A1

GENERAL INFORMATION:
 APPLICANT: Genon Corporation
 APPLICANT: Gregg, Morin B.
 APPLICANT: Walter, Funk D.
 APPLICANT: Mieczyslaw, Piatydzek A.
 TITLE OF INVENTION: A Second Mammalian Telomerase
 FILE REFERENCE: 080/003C
 CURRENT APPLICATION NUMBER: US/09/972,115A
 CURRENT FILING DATE: 2001-10-05
 PRIOR APPLICATION NUMBER: US 60/128,577
 PRIOR FILING DATE: 2000-04-10
 PRIOR APPLICATION NUMBER: US 60/129,123
 PRIOR FILING DATE: 1999-04-13
 NUMBER OF SEQ ID NOS: 64
 SOFTWARE: Patentin version 3.1

SEQ ID NO 8
 LENGTH: 1327
 TYPE: PR1
 ORGANISM: Homo sapiens
 US-09-972-115A-8

Query Match 11.7%; Score 90.5; DB 10; Length 1327;
 Best Local Similarity 28.4%; Pred. No. 2.9;
 Matches 42; Conservative 16; Mismatches 61; Indels 29; Gaps 6;

QY 6 VAAFAAI-VVSGSALAGVVPQWGGGNNHGGNSGSPDSTLSIYQGSANALALQSDAR 64
 DB 99 VAAAPVPAVSTSSAGVAPNPAGSGSNNSSPSSSSPTSS-SSSSPSSPGSSLAESPFAA 157

QY 65 KSEITIT-----OSGVNGADVQGDNSITIELTONGFRNNATIDQNNAKNYDQLYTVRYVT 120
 DB 158 GVSSTAPLPGAGAGCTGVPVAVSGALRELBACRNG-----DVSRRVRLVDA---- 204

QY 121 HEMAHANOTASD-----SSVMVROYVFG 143
 DB 205 -----ANVNAKDMAGRKSSPLHFAAGFG 227

RESULT 7
 US-10-199-937-4
 ; Sequence 4, Application US/10199937
 ; Publication No. US20030190739A1

GENERAL INFORMATION:
 APPLICANT: Christenson, Erik
 APPLICANT: Demaggio, Anthony J.
 APPLICANT: Goldman, Phyllis S.
 APPLICANT: McElligott, David L.
 TITLE OF INVENTION: TANKYRASE2 MATERIALS AND METHODS
 FILE REFERENCE: 27666/36559
 CURRENT APPLICATION NUMBER: US/10/199,937
 CURRENT FILING DATE: 2002-07-22
 PRIOR APPLICATION NUMBER: US/09/606,035
 PRIOR FILING DATE: 2000-06-28
 PRIOR APPLICATION NUMBER: 60/141,582
 PRIOR FILING DATE: 1999-06-29
 NUMBER OF SEQ ID NOS: 178
 SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 4
 LENGTH: 1327

RESULT 8
US-10-156-761-9343
Sequence 9343, Application US/10156761

Query Match	11.4%	Score 88.5;	DB 14,	Length 438;
Best Local Similarity	24.6%	Pred. NO. 1.1;		
Matches 34; Conservative	23;	Mismatches	46;	Indels 35; Gaps 6;

RESULT 9
US-10-289-762-171

```

: Sequence 177 Application US/10289762
: Publication No. US20040006218A1
:
: GENERAL INFORMATION:
: APPLICANT: Grifflats, R.
: TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
: TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention

```

Query Match	11.2%	Score 87	DB 15	Length 1156
Best Local Similarity	32.0%	Pred. NO. 5.5		
Matches 32	Conservative 14	Mismatches 34	Indels 20	Gaps 5

RESULT 10
US-10-156-761-8763
; Sequence 8763, Application US/10156761
; Publication No. US20030119018A1

Query Match	11.1%;	Score 86;	DB 14;	Length 482;
Best Local Similarity	30.2%;	Pred. No. 2.3;		
Matches 38;	Conservative 10;	Mismatches 58;	Indels 20;	Gaps 5

RESULT 11
US-10-156-761-11721

```

; Sequence 11721, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:

```

```

APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 11721
LENGTH: 271
TYPE: PRT
ORGANISM: Streptomyces avermitilis
US-10-156-761-11721

```

```

Query Match          11.0%; Score 85.5; DB 14; Length 271;
Best Local Similarity 25.8%; Pred. No. 1.2;
Matches 39; Conservative 18; Mismatches 43; Indels 51; Gaps 7;

```

```

QY 15 SCSALAG-----VWPQGGGNNHGGSSGPDSTLSIYQGSANALAL 59
DB 123 SCSAPGCGADPRAGGSPTPSSVSGGGGGGGGTSGSATL-----CGPALSV 176
QY 60 OSDARKS-----ETTQSGYGNAD-----VGO-GADNSTIETLTON----- 95
DB 177 STPEVREASDKRCEKVTVFHNTGGGAASGTVTFGTHIIGALGIDWATTESTEDLPVPI 236
QY 96 ---GFRNNAT-----IDQNAKNYDQLVTRVYT 120
DB 237 AAGATRKKTWYVCVDMARVPLGMHIEITRDVS 267

```

```

RESULT 12
US-09-880-748-1136
Sequence 1136, Application US/09880748
GENERAL INFORMATION:
APPLICANT: Ruben et al.
FILE REFERENCE: PF523
TITLE OF INVENTION: Antibodies that Immunospecifically Bind Blys
CURRENT APPLICATION NUMBER: US/09/880,748
CURRENT FILING DATE: 2001-06-15
PRIOR APPLICATION NUMBER: 60/212,210
PRIOR FILING DATE: 2000-06-15
PRIOR APPLICATION NUMBER: 60/240,816
PRIOR FILING DATE: 2000-10-17
PRIOR APPLICATION NUMBER: 60/276,248
PRIOR FILING DATE: 2001-03-16
PRIOR APPLICATION NUMBER: 60/277,379
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/293,499
PRIOR FILING DATE: 2001-05-25
NUMBER OF SEQ ID NOS: 3239
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1136
LENGTH: 254
TYPE: PRT
ORGANISM: Homo sapiens
US-09-880-748-1136

```

```

Query Match          10.8%; Score 84; DB 10; Length 254;
Best Local Similarity 24.5%; Pred. No. 1.6;
Matches 37; Conservative 19; Mismatches 61; Indels 34; Gaps 5;
QY 18 ALAGVVPWG-----GGNNHGGSSGPDSTLSIYQGSANALALOSDARKSE- 67

```

```

DB 106 ATTGALDMRGRLTVTVSSGGGGGGGGGGCG-----GSAQAVLTQPSVSCTPG 155
QY 68 TTITSGYNGADVGGAD-----NSTIETLTONGFRNNATIDQNAKNYDQLVTR 117
DB 156 QTITISCTGSDSNIGAGYVHWYQHLPCTAPKLLIYVNTNRPSGVDRFSGSKSTLASL 215
QY 118 VVT---HEMAHANQASDSSVMVTRQVGFGN 144
DB 216 VITGLQAEDEADYCCSYDLSLGRNYVFGS 246

```

```

RESULT 13
US-09-880-748-1165
Sequence 1165, Application US/09880748
GENERAL INFORMATION:
APPLICANT: Ruben et al.
TITLE OF INVENTION: Antibodies that Immunospecifically Bind Blys
FILE REFERENCE: PF523
CURRENT APPLICATION NUMBER: US/09/880,748
CURRENT FILING DATE: 2001-06-15
PRIOR APPLICATION NUMBER: 60/212,210
PRIOR FILING DATE: 2000-06-15
PRIOR APPLICATION NUMBER: 60/240,816
PRIOR FILING DATE: 2000-10-17
PRIOR APPLICATION NUMBER: 60/276,248
PRIOR FILING DATE: 2001-03-16
PRIOR APPLICATION NUMBER: 60/277,379
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/293,499
PRIOR FILING DATE: 2001-05-25
NUMBER OF SEQ ID NOS: 3239
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1165
LENGTH: 254
TYPE: PRT
ORGANISM: Homo sapiens
US-09-880-748-1165

```

```

Query Match          10.8%; Score 84; DB 10; Length 254;
Best Local Similarity 24.5%; Pred. No. 1.6;
Matches 37; Conservative 19; Mismatches 61; Indels 34; Gaps 5;

```

```

QY 18 ALAGVVPWG-----GGNNHGGSSGPDSTLSIYQGSANALALOSDARKSE- 67
DB 106 ATTGALDMRGRLTVTVSSGGGGGGGGGGCG-----GSAQAVLTQPSVSCTPG 155
QY 68 TTITSGYNGADVGGAD-----NSTIETLTONGFRNNATIDQNAKNYDQLVTR 117
DB 156 QTITISCTGSDSNIGAGYVHWYQHLPCTAPKLLIYVNTNRPSGVDRFSGSKSTLASL 215
QY 118 VVT---HEMAHANQASDSSVMVTRQVGFGN 144
DB 216 VITGLQAEDEADYCCSYDLSLGRNYVFGS 246

```

```

RESULT 14
US-10-156-761-13168
Sequence 13168, Application US/10156761
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089

```

; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 13168
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-13168

Query Match 10.8%; Score 83.5; DB 14; Length 350;
Best Local Similarity 24.6%; Pred. No. 2.8;
Matches 43; Conservative 19; Mismatches 76; Indels 37; Gaps 6;

QY 1 MKLKVAFAFAIV-----SSALAGVVPQMGCGN--HNGGG 36
DB 6 LRLAVALTAVALPLSLATPTASADTCVKSRRPSGKVLQGYWENWDSSNGVHPFG 65
QY 37 NSSGPDSTLSIYQGSANALALQSDARKSETTITQSGYNGADVGQAD-----NST 89
DB 66 WTPITDSRIAHGYNVINAAPV---IRSDGTALMEDGMDTGAVKATPAEMCQAKASQT 122
QY 90 IELTQNGFRNATIDQWNAKNYDQVTRVVTHEMAHANQIASDSSVMVROYFGN 144
DB 123 ILISIGAA--AGIDLSAVADRFAVATVP-LIKKYNFDGIDIDIEGTGLVSGN 174

RESULT 15

US-09-952-267-15
; Sequence 15, Application US/09952267
; Publication No. US20030032772A1
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AERI, CHRISTOPH
; APPLICANT: CORE, LESLIE D.
; APPLICANT: MACTIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USPA1 AND USPA2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY.024
; CURRENT APPLICATION NUMBER: US/09/952,267
; CURRENT FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: 09/336,447
; PRIOR FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
; LENGTH: 889
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-952-267-15

Query Match 10.7%; Score 83; DB 10; Length 889;
Best Local Similarity 27.9%; Pred. No. 10;
Matches 39; Conservative 16; Mismatches 51; Indels 34; Gaps 8;

QY 28 GGGNHNGGSSGPDSTLSIYQ---YGSANALALQSDARKSETTITQSGY-----NGGA 79
DB 69 GAGRHNVGSGSAHSGILGCKNTVNGYTSIIVGYNETQGDITFVGGYKNLAKGNYT 128
QY 80 DVGG-----ADNSTIELTQNGFRN-----NATID---QWNAKNYDQVTRVTHEMAH 125
DB 129 FVGGGYKNLAEGDNAT---AGGFANLAEGDNATAGFENRABGIDSVSG-----GY 179
QY 126 ANQTASDSSVMVROYFGN 145
DB 180 ANQATGSSSTV---AGGSNN 196

Search completed: March 11, 2004, 19:18:39
Job time : 25.6 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:24:14 ; Search time 10.3 Seconds
(without alignments)
1410.186 Million cell updates/sec

Title: US-09-543-407-24

Perfect score: 775

Sequence: 1 MLLKLVAAFAAIVSGSALA.....DSSVMVROYFGNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	690	89.0	151	2 JCG639	fimbria protein ag
2	690	89.0	151	2 A10635	major curlin chain
3	533	68.8	151	2 S70788	curlin protein csg
4	511.5	66.0	152	2 D90806	curlin major subun
5	511.5	66.0	152	2 H85665	hypothetical prote
6	101	13.0	409	2 T20847	hypothetical prote
7	99	12.8	1748	2 S42136	cnjB protein - Tet
8	98.5	12.7	151	2 JCG640	fimbria protein ag
9	98.5	12.7	151	2 AH0635	nucleation compone
10	96.5	12.5	151	2 S70787	curlin nucleator p
11	96.5	12.5	151	2 C90806	minor curlin subun
12	96.5	12.5	151	2 G85665	curlin minor chain
13	93.5	12.1	440	2 D70604	probable PPE prote
14	92.5	11.9	440	2 AD1539	probable sugar ABC
15	92.5	11.9	1028	2 A56038	DNA-binding protei
16	92.5	11.9	1213	2 S16356	ovo protein - FruI
17	90	11.6	145	2 AD3143	conserved hypotet
18	90	11.6	145	2 H98144	hypothetical prote
19	89.5	11.6	347	2 B39112	merozoite 45K surf
20	89.5	11.5	256	2 T03371	glycine-rich prote
21	89	11.5	262	2 S00375	tail fiber protein
22	89	11.5	573	2 C86266	F3F19.21 protein -
23	89	11.5	2174	2 E95965	hypothetical glyci
24	88.5	11.4	321	2 A47369	RNA-binding protei
25	87	11.2	590	1 A45621	leishmanolysin (EC
26	87	11.2	590	1 B42049	leishmanolysin (EC
27	87	11.2	599	2 A44951	leishmanolysin (EC
28	87	11.2	602	1 P10221	leishmanolysin (EC
29	87	11.2	646	1 S19916	leishmanolysin (EC

30	87	11.2	1537	2 F86509	CT147 hypothetical
31	87	11.2	1537	2 C81558	conserved hypotet
32	86	11.1	447	2 G84687	probable disease r
33	86	11.1	582	2 F70675	probable PPE prote
34	85.5	11.0	438	2 T35789	probable secreted
35	85	11.0	401	2 C88571	protein C05B5.3 [1
36	85	11.0	575	2 S35327	protein kinase agg
37	85	11.0	967	2 S66852	hypothetical prote
38	84.5	10.9	340	2 A83401	hypothetical prote
39	84	10.8	439	2 AC1182	probable sugar ABC
40	84	10.8	764	2 H71607	hypothetical prote
41	83.5	10.8	423	2 T19581	hypothetical prote
42	83.5	10.8	1635	2 A10452	hemolysin (importe
43	83	10.7	639	2 C42049	leishmanolysin (EC
44	82.5	10.6	343	2 T05221	hypothetical prote
45	82	10.6	382	2 F90892	probable outer mem

ALIGNMENTS

RESULT 1

JCG639 fimbria protein agfa precursor - Salmonella enteritidis

C:Species: Salmonella enteritidis

C:Date: 31-Dec-1996 #sequence revision 31-Dec-1996 #text_change 08-Oct-1999

C:Accession: JCG639; PC6015; A44898

R:Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Baner, P.A.; Kay, W.W.

J. Bacteriol. 178, 662-667, 1996

A:Title: Salmonella enteritidis agfaBC operon encoding thin, aggregative fimbriae.

A:Reference number: JCG639; MUID:96146512; PMID:8550497

A:Accession: JCG639

A:Molecule type: DNA

A:Residues: 1-151 <COL>

A:Cross-references: GB:U43280; NID:91184712; PIDN:AA04359.1; PID:91184714

A:Accession: PC6015

A:Molecule type: protein

A:Residues: 21-52 <CO2>

A:Experimental source: strain 27655-3b

A>Note: the authors translated the codon ACG for residue 44 as Ile

R:Collinson, S.K.; Emdy, L.; Muller, K.H.; Trust, T.J.; Kay, W.W.

J. Bacteriol. 173, 4773-4781, 1991

A:Title: Purification and characterization of thin, aggregative fimbriae from Salmonella

A:Reference number: A44898; MUID:91310586; PMID:1677357

A:Accession: 27655

A:Contents: A44898

A:Status: preliminary

A:Molecule type: protein

A:Residues: 21-33 <CO3>

A>Note: sequence extracted from NCBI backbone (NCBIP:45936)

C:Genetics:

A:Gene: agfa

C:Function:

A:Description: major component of thin aggregative fimbriae

A>Note: fimbriae bind to fibronectin, plasminogen, tissue plasminogen activator

C:Keywords: fimbria

F:1-20/Domain: signal sequence #status predicted <Sig>

F:21-151/Product: fimbria protein agfa #status experimental <Mat>

Query Match Best Local Similarity 89.0%; Score 690; DB 2; Length 151; Pred. No. 4, 7e-51; Matches 138; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

QY	1	MLLKLVAAFAAIVSGSALAGVPPMGGGGNNNGSGSPDSTSIYOGSNNALALQ	60
DB	1	MLLKLVAAFAAIVSGSALAGVPPMGGGGNNNGSGSPDSTSIYOGSNNALALQ	60
QY	61	SDARKSETTITSGYGNGADVGQADNSFTIELTONGFRNNATIDQWANNYDQLYRVVT	120
DB	61	SDARKSETTITSGYGNGADVGQADNSFTIELTONGFRNNATIDQWANNYDQLYRVVT	120
QY	121	HEMAHANTQASDSSVMVROYFGNNATANQY 151	
DB	121	HEMAHANTQASDSSVMVROYFGNNATANQY 151	

Db 121 NNAALVNOTASDSSVMVROVGFNNATANY 151

RESULT 2

Major curlin chain precursor (imported) - Salmonella enterica subsp. enterica serovar Typhimurium

C/Species: Salmonella enterica subsp. enterica serovar Typhimurium

A/Note: This species has also been called Salmonella typhimurium

C/Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002

C/Accession: A10635

R/Parthill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Mowle, S.; O'Gaora, P.

A/Author: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.

A/Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhimurium

A/Reference number: A80502; MUID:21534947; PMID:11677608

A/Accession: A10635

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-151 <PAR>

A/Cross-references: GB:AL513382; PIDN:CAD08268.1; PID:G16502315; GSPDB:GN00176

C/Genetics:

A/Gene: STY1181

Query Match 89.0%; Score 690; DB 2; Length 151;

Best Local Similarity 91.4%; Pred. No. 4.7e-51;

Matches 138; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

Db 1 MKLKVAFPAIVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYOGSANAALALQ 60

1 MKLKVAFPAIVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYOGSANAALALQ 60

Qy 61 SDARKSETTITGSGYNGADVGGAGDNSTIELTONGFNNATIDOMNAKXVDQVTRVVT 120

Db 61 SDARKSETTITGSGYNGADVGGAGDNSTIELTONGFNNATIDOMNAKXVDQVTRVVT 120

Qy 121 HEMAHANOTASDSSVMVROVGFNNATANY 151

121 HEMAHANOTASDSSVMVROVGFNNATANY 151

Db 121 NNAALVNOTASDSSVMVROVGFNNATANY 151

RESULT 3

70788

Curlin protein csgA precursor - Escherichia coli (strain K-12)

N/Alternate names: csgA protein; major curlin protein

C/Species: Escherichia coli

C/Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 01-Mar-2002

C/Accession: S70788; G64846; S31202; S34560; S34559

R/Hammar, M.; Arngvist, A.; Bijn, Z.; Olsen, A.; Normark, S.

Mol. Microbiol. 18, 661-670, 1995

A/Title: Expression of two csg operons is required for production of fibronectin- and C

A/Reference number: S70783; MUID:96414468; PMID:8817489

A/Accession: S70788

A/Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-151 <HAM>

A/Cross-references: EMBL:X90754; NID:g147558; PIDN:CAA62282.1; PID:g147564

A/Experimental source: strain K12, substrain W3110

A/Note: The nucleotide sequence was submitted to the EMBL Data Library, August 1995

R/Battner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co

A/ Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A/Title: The complete genome sequence of Escherichia coli K-12.

A/Reference number: A64720; MUID:97426617; PMID:9278503

A/Accession: G64846

A/Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-151 <BAT>

A/Cross-references: GB:AE000205; GB:U00096; NID:g1787265; PIDN:AACT4126.1; PID:g1787279;

A/Experimental source: strain K-12, substrain MG1655

R/Olsen, A.; Arngvist, A.; Hammar, M.; Sukupolvi, S.; Normark, S.

Mol. Microbiol. 7, 523-536, 1993

A/Title: The Rpos sigma factor relieves H-NS-mediated transcriptional repression of csg

A/Reference number: S31202; MUID:93211294; PMID:8459772

A/Accession: S31202

A/Molecule type: DNA

A/Residues: 1-6, 'V', 8-151 <OLS1>

A/Cross-references: EMBL:L04979

A/Accession: S34560

A/Molecule type: protein

A/Residues: 21-42/44-50 <OLS2>

R/Olsen, A.N.; Arngvist, A.M.

Submitted to the EMBL Data Library, October 1992

A/Reference number: S34559

A/Accession: S34559

A/Molecule type: DNA

A/Residues: 1-133, 'RORDSGMW', <OLS3>

A/Cross-references: EMBL:L04979; NID:g290424; PIDN:AAA23616.1; PID:g290425

A/Experimental source: strain K-12, substrain W3110

C/Genetics:

A/Gene: csgA

A/Map position: 23.15

C/Function:

A/Description: major component of wild-type curli; interaction between CsgA and CsgB tr

A/Note: curli are thin, coiled fibers expressed on the surface of Escherichia coli that

F/1-20/Domain: signal sequence #status predicted <Sig>

F/21-151/Product: curlin #status experimental <Mat>

Query Match 68.8%; Score 533; DB 2; Length 151;

Best Local Similarity 70.2%; Pred. No. 7.1e-38;

Matches 106; Conservative 18; Mismatches 27; Indels 0; Gaps 0;

Db 1 MKLKVAFPAIVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYOGSANAALALQ 60

1 MKLKVAFPAIVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYOGSANAALALQ 60

Qy 61 SDARKSETTITGSGYNGADVGGAGDNSTIELTONGFNNATIDOMNAKXVDQVTRVVT 120

Db 61 SDARKSETTITGSGYNGADVGGAGDNSTIELTONGFNNATIDOMNAKXVDQVTRVVT 120

Qy 121 HEMAHANOTASDSSVMVROVGFNNATANY 151

121 HEMAHANOTASDSSVMVROVGFNNATANY 151

Db 121 GNGAIVDQTSNNSVNTQVGFNNATANY 151

RESULT 4

90806

Curlin major subunit CsgA (imported) - Escherichia coli (strain O157:H7, substrain R1MD

C/Species: Escherichia coli

C/Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001

C/Accession: D90806

R/Hayashi, T.; Makino, K.; Onishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G

Gawawara, N.; Yasunaga, T.; Kuhara, S.; Shibata, T.; Hattori, M.; Shinagawa, H.

Genome Res. 8, 11-22, 2001

A/Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen

A/Reference number: A99629; MUID:21156231; PMID:11258796

A/Accession: D90806

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-152 <HAY>

A/Cross-references: GB:BA000007; PIDN:BA034843.1; PID:g13360880; GSPDB:GN00154

A/Experimental source: strain O157:H7, substrain R1MD 0509952

C/Genetics:

A/Gene: ECS1420

Query Match 66.0%; Score 511.5; DB 2; Length 152;

Best Local Similarity 68.4%; Pred. No. 4.6e-36;

Matches 104; Conservative 19; Mismatches 28; Indels 1; Gaps 1;

Qy 1 MKLKVAFPAIVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYOGSANAALALQ 59

1 MKLKVAFPAIVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYOGSANAALALQ 60

Db 1 MKLKVAFPAIVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYOGSANAALALQ 60

Qy 60 QSDARKSETTITGSGYNGADVGGAGDNSTIELTONGFNNATIDOMNAKXVDQVTRVVT 119

Db 61 QADARSDLTITQHGCGNADVGCGSDSSIDLTFQGFNSATLTDQMGKDSMTVKQFG 120
 QY 120 THEMAHANTASDSVWVRQVFGNNATNAY 151
 Db 121 GGNGAIVDOTASNSTVNTVQVFGNNATNAY 152

RESULT 5

H5665
 Hypothetical protein csgA [imported] - Escherichia coli (strain O157:H7, substrain EDL93
 C:Species: Escherichia coli
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 C:Accession: H85665
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.U.; Mayhew
 Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
 Nature 409, 529-533, 2001
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A:Reference number: A85480; MUID:21074935; PMID:11206551
 A:Accession: H85665
 A:Status: preliminary
 A:Molecule type: DNA

A:Residues: 1-152 <STO>

A:Cross-references: GB:AE005174; NID:g12514574; PIDN:RAG55788.1; GSPDB:GN00145; UMGD:216
 A:Experimental source: strain O157:H7, substrain EDL933
 C:Genetics:

A:Gene: csgA

Query Match 66.0%; Score 511.5; DB 2; Length 152;
 Best Local Similarity 68.4%; Pred. No. 4.6e-36;
 Matches 104; Conservative 19; Mismatches 28; Indels 1; Gaps 1;

QY 1 MKLLKVAAPAAIYVSSGALAGVPPW-GGGGNHNGGNSGPDSTLSIYQGSANALAL 59
 Db 1 MLILKVAIAIYVSSGALAGVPPWYCGGGNHGGGNSGPDSELTITQYGGGALAL 60

QY 60 QSDARKSETTITQSGYGNAGDVCGADNSTIELTQNGFNNAITIDQMAKNTDQVTRVV 119
 Db 61 QADARSDLTITQHGCGNADVGCGSDSSIDLTFQGFNSATLTDQMGKDSMTVKQFG 120

QY 120 THEMAHANTASDSVWVRQVFGNNATNAY 151
 Db 121 GGNGAIVDOTASNSTVNTVQVFGNNATNAY 152

RESULT 6

T20847
 Hypothetical protein F13B9.4 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 02-Aug-2002

C:Accession: T20847

R:McMurray, A.
 submitted to the EMBL Data Library, February 1996

A:Reference number: Z19332

A:Accession: T20847

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-409 <MIL>

A:Cross-references: EMBL:Z69383; PIDN:CAA93412.1; GSPDB:GN00022; CESP:F13B9.4

A:Experimental source: clone F13B9

C:Genetics:

A:Gene: CESP:F13B9.4

A:Map position: 4

A:Introns: 32/1; 275/3; 337/3

C:Superfamily: Iorictin

Query Match 13.0%; Score 101; DB 2; Length 409;
 Best Local Similarity 23.3%; Pred. No. 0.4;
 Matches 37; Conservative 26; Mismatches 66; Indels 30; Gaps 5;

QY 15 SGGALAGVVPWCGGNGHNGGNSGPDSTLSIYQGSANALALQSDARKSETTITQSG 74
 Db 134 SGGAGSGMNSFGGCGYGNONGFGGSGFSGSGMGSNSLSANSNGNNG--SSSG 191

QY 75 YGN--GADVCGADNSTIE--LTONGFRNN-----ATTQMAK 109
 Db 192 YQNNQGRHQGGCGGSSSSSVMSNNGYSSNGYGNNGNPTPSFLNNVSSAAQDYINIV 251
 QY 110 NYDQVTRVYTHEMAHANTASDSVWVRQVFGNNATA 148
 Db 252 NNKSLITVQINIQ--ASNWASNSVQAQYIQYETNRSA 287

RESULT 7

S42136

cnjB protein - Tetrahymena thermophila

C:Species: Tetrahymena thermophila

C:Date: 19-Mar-1997 #sequence_revision 18-Jul-1997 #text_change 07-Dec-1999

C:Accession: S42136; S42135; S03650

R:Taylor, F.M.; Martindale, D.W.
 submitted to the EMBL Data Library, October 1992

A:Reference number: S42136

A:Accession: S42136

A:Molecule type: DNA

A:Residues: 1-1748 <TAY>

A:Cross-references: EMBL:L03710; NID:g161751; PID:g161752

R:Taylor, F.M.; Martindale, D.W.

Nucleic Acids Res. 21, 4610-4614, 1993

A:Title: Retroviral-type zinc fingers and glycine-rich repeats in a protein encoded by

A:Reference number: S42135; MUID:94051569; PMID:8233798

A:Accession: S42135

A:Molecule type: DNA

A:Residues: 1164-1174; 1179-1198; 1233-1252; 1285-1293; 1297-1309; 1316-1326; 1331-1341; 1343-

A:Cross-references: EMBL:L03710

R:Martindale, D.W.; Taylor, F.M.

Nucleic Acids Res. 16, 2189-2201, 1988

A:Title: Multiple introns in a conjugation-specific gene from Tetrahymena thermophila.

A:Reference number: S03650; MUID:86189811; PMID:3357771

A:Accession: S03650

A:Molecule type: DNA

A:Residues: 236-250; 'I', 252-255; 'N', 257-773 <MAR>

A:Cross-references: EMBL:X06462

C:Genetics:

A:Gene: cnjB

A:Genetic code: SGC5

A:Introns: 85/3; 136/1; 157/3; 201/2; 290/2; 327/3; 499/1; 573/2; 607/3; 708/3; 777/3; 8

C:Keywords: zinc finger

F:1164-1450/Region: glycine-rich

F:1451-1464/Region: zinc finger CCHC motif

F:1478-1491/Region: zinc finger CCHC motif

F:1501-1514/Region: zinc finger CCHC motif

F:1530-1543/Region: zinc finger CCHC motif

F:1555-1568/Region: zinc finger CCHC motif

F:1579-1592/Region: zinc finger CCHC motif

F:1602-1615/Region: zinc finger CCHC motif

F:1626-1748/Region: glycine-rich

Query Match 12.8%; Score 99; DB 2; Length 1748;
 Best Local Similarity 28.0%; Pred. No. 3.1;
 Matches 37; Conservative 16; Mismatches 31; Indels 48; Gaps 6;

QY 25 QWGGGNGHNG--GNSGPDSTLSIYQGSANALALQSDARKSETTIT--QSGYXN 77
 Db 1640 QFGGGSNSNGSGSWGSSGSDWN-----COSNVQESTTSSGGGSGSGSN 1685

QY 78 GADVCGADNSTIELTQNGFRNNATIDQMAKNTDQVTRVYTHEMAHANTASDSVWV 137
 Db 1686 QTGGGKSGNDN-----QQQNTNTGGGKSGSN-----SNTNTNSS-- 1722

QY 138 RQVFGNNATNAN 149
 Db 1723 ----WGSNNQAS 1730

RESULT 8

JC6040

fimbria protein agfB precursor - *Salmonella enteritidis*
 C:Species: *Salmonella enteritidis*
 C:Date: 31-Dec-1996 #sequence_revision 31-Dec-1996 #text_change 08-Oct-1999
 C:Accession: J06040
 R:Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Bannery, P.A.; Kay, W.W.
 J. Bacteriol. 178, 662-667, 1996
 A:Title: *Salmonella enteritidis* agfBAC operon encoding thin, aggregative fimbriae.
 A:Reference number: J06039; MUID:96146512; PMID:8550497
 A:Accession: J06040
 A:Molecule type: DNA
 A:Residues: 1-151 <COL>
 A:Cross-references: GB:U043280; NID:g1184712; PID:AA043598.1; PID:g1184713
 A:Experimental source: strain 276755-3b
 C:Genetics:
 A:Gene: agfB
 C:Function:
 A:Description: minor component of thin aggregative fimbriae
 A:Note: fimbriae bind to fibronectin, plasminogen, tissue plasminogen activator
 C:Keywords: fimbria
 F:1-21/Domain: signal sequence #status predicted <SIG>
 F:12-151/Product: fimbria protein agfB #status predicted <MAT>

Query Match 12.7%; Score 98.5; DB 2; Length 151;
 Best Local Similarity 28.8%; Pred. No. 0.21;
 Matches 34; Conservative 17; Mismatches 48; Indels 19; Gaps 5;

QY 51 GSANAAALQSDARKSE-----TTTOSGYNGADVGO-GADNST-----IELTQ 94
 DB 14 GAGGATATNTVDLARSFNPVNELSKSFNOAIIQGVTDNSARRROESKLSVTSQ 73
 QY 95 NGFRNNATIDQWNAKNTD-QLVTRVVTHEMAHNOTASDSSVWVRQVFGNNATANQY 151
 DB 74 EGGNNRAKVQD--AGNYNFAVIEOTGNANDASISQSYNGNSAIIQKSGSKNANTQY 129

RESULT 9

AH0635
 nucleation component of curlin monomers [imported] - *Salmonella enterica* subsp. *enterica*
 C:Species: *Salmonella enterica* subsp. *enterica* serovar Typhi
 A:Note: this species has also been called *Salmonella typhi*
 C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
 C:Accession: AH0635
 R:Perkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, J.; Moule, S.; O'Gaora, P.
 Nature 413, 848-852, 2001
 A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
 A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serov.
 A:Reference number: AB0502; MUID:21534947; PMID:11677608
 A:Accession: AH0635
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-151 <PAR>
 A:Cross-references: GB:AL513382; PIDN:CAD08267.1; PID:g15052314; GSPDB:GN00176
 C:Genetics:
 A:Gene: STY1180

Query Match 12.7%; Score 98.5; DB 2; Length 151;
 Best Local Similarity 28.8%; Pred. No. 0.21;
 Matches 34; Conservative 17; Mismatches 48; Indels 19; Gaps 5;

QY 51 GSANAAALQSDARKSE-----TTTOSGYNGADVGO-GADNST-----IELTQ 94
 DB 14 GAGGATATNTVDLARSFNPVNELSKSFNOAIIQGVTDNSARRROESKLSVTSQ 73
 QY 95 NGFRNNATIDQWNAKNTD-QLVTRVVTHEMAHNOTASDSSVWVRQVFGNNATANQY 151
 DB 74 EGGNNRAKVQD--AGNYNFAVIEOTGNANDASISQSYNGNSAIIQKSGSKNANTQY 129

RESULT 10
 S70787
 curlin nucleator protein csgB precursor - *Escherichia coli* (strain K-12)

N:Alternate names: csgB protein; curlin nucleation component; minor curlin protein
 C:Species: *Escherichia coli*
 C:Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 01-Mar-2002
 C:Accession: S70787; F64846
 R:Hammer, M.; Arngvist, A.; Bian, Z.; Olsen, A.; Normark, S.
 Mol. Microbiol. 18, 661-670, 1995
 A:Title: Expression of two csg operons is required for production of fibronectin- and C.
 A:Reference number: S70783; MUID:96414468; PMID:8617489
 A:Accession: S70787
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-151 <HAM>
 A:Cross-references: EMBL:X90754; NID:g1147558; PIDN:CAA62281.1; PID:g1147563
 A:Experimental source: strain K12, substrain W3110
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C.
 A.; Rose, D.; Mau, B.; Shaoh, Y.
 Science 277, 1453-1462, 1997
 A:Title: The complete genome sequence of *Escherichia coli* K-12.
 A:Reference number: A64720; MUID:97426617; PMID:9278503
 A:Accession: F64846
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-151 <BLAT>
 A:Cross-references: GB:AE000205; GB:U00096; NID:g1787265; PIDN:AA074125.1; PID:g1787278
 A:Experimental source: strain K-12, substrain MG1655
 C:Genetics:
 A:Gene: csgB
 A:Map position: 23.15
 C:Function:
 A:Description: minor component of wild-type curli; interaction between CsgA and CsgB tr
 A:Note: curli are thin, coiled fibers expressed on the surface of *Escherichia coli* that
 F:1-21/Domain: signal sequence #status predicted <SIG>
 F:22-151/Product: minor curlin chain #status predicted <MAT>

Query Match 12.5%; Score 96.5; DB 2; Length 151;
 Best Local Similarity 25.6%; Pred. No. 0.31;
 Matches 30; Conservative 24; Mismatches 46; Indels 17; Gaps 4;

QY 51 GSANAAALQSDARKSE-----TTTOSGYNGADVGO-GADNSTIELTQNGFRNNATIDQ 105
 DB 14 GAGGIAAAGYDLANSEVNPVNELSKSFNOAIIQAGTNNASQALQSGSKLAIVAAQ 73
 QY 106 WNAKYVDQLVTRVVTHEMAHANO--TASDSSV-----MVRVGFNNATANQY 151
 DB 74 EGSSNRK- IDQTDYNTLVITDQAGSANDASISQAYGNTAVITQKSGSKNANTQY 129

RESULT 11

C90806
 minor curlin subunit precursor CsgB [imported] - *Escherichia coli* (strain O157:H7, subs
 C:Species: *Escherichia coli*
 C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
 C:Accession: C90806
 R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G
 gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 8, 11-22, 2001
 A:Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and gen
 A:Reference number: A99629; MUID:21156231; PMID:11258796
 A:Accession: C90806
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-151 <HAY>
 A:Cross-references: GB:BA000007; PIDN:BA034842.1; PID:g13360879; GSPDB:GN00154
 A:Experimental source: strain O157:H7, substrain R1MD 0509952
 C:Genetics:
 A:Gene: EC01419

Query Match 12.5%; Score 96.5; DB 2; Length 151;
 Best Local Similarity 25.6%; Pred. No. 0.31;
 Matches 30; Conservative 24; Mismatches 46; Indels 17; Gaps 4;

Db	59	LONAAAAYIMASGSG-----GGGCTGNGGGGASGPGGSPSANSGGGGGG-----	104
Qy	63	ARKSETTITGSGYGADVGAGADNSTIELTONGFRNATIDOMAKNYDQVTRVYTHE	122
Db	105	-----GGNGYINC GGVG--GPNN--LDGNNLILFASVSNYNESN-----SKFHNNH	147
Qy	123	MAHA-----NOTASDSSVVRQVFGNNATA	148
Db	148	HHHCHNNNNNNNGGOTSMGHPFYGNPSA	177

Search completed: March 11, 2004, 18:42:11
 Job time : 11.3 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:14:48 ; Search time 6.3 Seconds

(without alignments)
1248.031 Million cell updates/sec

Title: US-09-543-407-24

Perfect score: 775

Sequence: 1 MKLLRVAAFPALIVSGSALA.....DSSVMVROYGNNATANYQ 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	690	89.0	151	1	CSGA_SALTY
2	533	68.8	151	1	CSGA_ECOLI
3	511.5	66.0	151	1	CSGA_ECO57
4	98.5	12.7	151	1	CSGB_SALTY
5	98.5	12.7	151	1	CSGB_SALTY
6	96.5	12.5	151	1	CSGB_ECOLI
7	92.5	11.9	1028	1	OVO_DROME
8	90.5	11.7	1327	1	TNKT_HUMAN
9	90	11.6	347	1	MSA2_PLA2
10	89	11.5	262	1	VG38_BPT2
11	87	11.2	590	1	GP63_LEIDO
12	87	11.2	599	1	GP63_LEICH
13	87	11.2	602	1	GP63_LEIMA
14	87	11.2	646	1	GP63_LEIME
15	87	11.0	1656	1	OMP_RICUA
16	85	11.0	401	1	YF48_MYCTU
17	82	10.6	678	1	PER_DROMI
18	82	10.6	1093	1	PER_DROMI
19	82	10.6	1115	1	TBC3_CHURE
20	81.5	10.5	306	1	HMXD_DROPS
21	81.5	10.5	311	1	7B4C_PSESP
22	81.5	10.5	392	1	HME1_HUMAN
23	81	10.5	165	1	GRPI_ORYSA
24	80.5	10.4	548	1	CEAK_ECOLI
25	80.5	10.4	720	1	GZAC_BREDI
26	80.5	10.4	894	1	ILF3_HUMAN
27	80	10.3	427	1	CP1A_DROME
28	80	10.3	491	1	YK98_MYCTU
29	79.5	10.3	172	1	CH18_DROME
30	79.5	10.3	760	1	YBIL_ECOLI
31	79.5	10.3	1034	1	ICEN_PANAN
32	79.5	10.3	1258	1	ICEN_ERHNE
33	79.5	10.3	1322	1	ICEN_PANAN

34	79	10.2	878	1	FIND_ECOLI	P30130 escherichia
35	79	10.2	933	1	NP3_HUMAN	Q81XFO homo sapien
36	78.5	10.1	718	1	FLGE_HELPY	Q92XY0 helicobacte
37	78.5	10.1	718	1	FLGE_HELPY	P50610 helicobacte
38	78	10.1	1067	1	SGG_DROME	P18431 drosophila
39	77	9.9	796	1	MBN_DROME	P52302 drosophila
40	77	9.9	1146	1	YHC3_YEAST	P38742 saccharomyc
41	77	9.9	1690	1	KF1A_HUMAN	Q12756 homo sapien
42	76.5	9.9	348	1	Y217_AQUAS	O66408 aquifex aeo
43	76.5	9.9	979	1	RFX1_HUMAN	P22670 homo sapien
44	76.5	9.9	1196	1	ICEV_PSESP	O33479 pseudomonas
45	76.5	9.9	1210	1	ICEN_PSEFL	P09815 pseudomonas

ALIGNMENTS

RESULT 1	CSGA_SALTY	STANDARD;	PRT;	151 AA.
ID	CSGA_SALTY			
AC	P55225;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Major curilin subunit precursor (Fimbriin SEF17).			
GN	CSGA OR AGFA OR STM144 OR STY1181 OR T1776.			
OS	Salmonella typhimurium,			
OS	Salmonella enteritidis.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OC	Enterobacteriaceae; Salmonella.			
OX	NCHI_TaxID=602; 601, 592;			
NP	[1]			
RC	SEQUENCE FROM N.A.			
RC	SPECIES=S.typhimurium; STRAIN=SR-11;			
RA	MEDLINE=98117058; PubMed=9457880;			
RT	Romling U., Bian Z., Hammar M., Sierraalta W.D., Normark S.;			
RT	"Curli fibers are highly conserved between Salmonella typhimurium and			
RT	Escherichia coli with respect to operon structure and regulation.";			
RL	J. Bacteriol. 180:722-731(1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;			
RC	MEDLINE=21534948; PubMed=11677609;			
RA	McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,			
RA	Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,			
RA	Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,			
RA	Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,			
RA	Waterston R., Wilson R.K.;			
RT	"Complete genome sequence of Salmonella enterica serovar Typhimurium			
RT	LT2.";			
RL	Nature 413:852-856(2001).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=S.typhimurium; STRAIN=CT18;			
RC	MEDLINE=21534947; PubMed=11677608;			
RA	Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,			
RA	Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,			
RA	Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,			
RA	Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrer J.,			
RA	Feltwell T., Hamlin N., Haque A., Hien T.T., Holtroyd S., Jagels K.,			
RA	Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,			
RA	Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,			
RA	Whitehead S., Barrall B.G.;			
RT	"Complete genome sequence of a multiple drug resistant Salmonella			
RT	enterica serovar Typhimurium CT18.";			
RL	Nature 413:848-852(2001).			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=S.typhimurium; STRAIN=Ty2 / ATCC 700931;			
RC	MEDLINE=22531167; PubMed=12644504;			
RA	Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,			
RA	Burford V., Kodoyianni V., Schwartz D.C., Blattner F.R.;			

RT "Comparative genomics of *Salmonella enterica* serovar Typhimurium strains Ty2
 RL and CT18.";
 RT J. Bacteriol. 185:2330-2337(2003).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC SPECIES-S. enteritidis; STRAIN=27655-3B;
 RA MEDLINE=96146512; PubMed=8550497;
 RX Collinson S.K., Cloughier S.C., Doran J.L., Baner P.A., Kay W.W.;
 RT "Salmonella enteritidis agfBAC operon encoding thin, aggregative
 RT fimbriae.";
 RL J. Bacteriol. 178:662-667(1996).
 RN [6]
 RP SEQUENCE OF 21-151 FROM N.A.
 RC SPECIES-S. enteritidis; STRAIN=27655-3B;
 RA MEDLINE=94013373; PubMed=8104955;
 RX Doran J.L., Collinson S.K., Burian J., Santos G., Todd E.C.D.,
 RA Munro C.K., Kay C.M., Baner P.A., Peterkin P.I., Kay W.W.;
 RT "DNA-based diagnostic tests for *Salmonella* species targeting agfa,
 RT the structural gene for thin, aggregative fimbriae.";
 RL J. Clin. Microbiol. 31:2263-2273(1993).
 RN [7]
 RP SEQUENCE OF 21-33.
 RC SPECIES-S. enteritidis; STRAIN=27655-3B;
 RX MEDLINE=91310586; PubMed=1677357;
 RA Collinson S.K., Emeody L., Mueller K.-M., Trust T.J., Kay W.W.;
 RT "Purification and characterization of thin, aggregative fimbriae from
 RT *Salmonella enteritidis*.";
 RL J. Bacteriol. 173:4773-4781(1991).
 CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 CC COLLED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 CC FIBRONECTIN.
 CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.jdb-eb.ch/announce/>
 CC or send an email to license@jdb-eb.ch).
 CC -----
 DR EMBL; AJ002301; CAA05317.1; -
 DR EMBL; AE008749; AAL20074.1; -
 DR EMBL; AE01627269; CAD08268.1; -
 DR EMBL; AE016840; AAC06399.1; -
 DR EMBL; U43280; AAC43599.1; -
 DR PIR; JC6039; JCG6039
 DR StyGene; SGI0608; csgA.
 KM Fimbria; Signal; Complete proteome.
 FT SIGNAL 1 20
 FT CHAIN 21 151 MAJOR CURLIN SUBUNIT.
 FT CONFLICT 134 151 SVMRYOYFGNNATANY -> DSYTYVAS (IN
 FT REF 6).
 SQ SEQUENCE 151 AA; 15305 MW; B7DACDD16B621359 CRC64;
 Query Match 89.0%; Score 690; DB 1; Length 151;
 Best Local Similarity 91.4%; Pred. No. 1.4e-51;
 Matches 138; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

RESULT 2
 ID CSGA_ECOLI STANDARD; PRT; 151 AA.
 AC P28307;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Major curlin subunit precursor.
 GN CSGA OR B1042.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / W3110;
 RX MEDLINE=93211294; PubMed=8459772;
 RA Olsen A., Argyris A.;
 RT "The Rpos sigma factor relieves H-NS-mediated transcriptional
 RT repression of csgA, the subunit gene of fimbriectin-binding curli in
 RT *Escherichia coli*.";
 RL Mol. Microbiol. 7:523-536(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MC4100;
 RX MEDLINE=96414468; PubMed=8817489;
 RA Hammer M., Argyris A., Bian Z., Olsen A., Normark S.;
 RT "Expression of two csg operons is required for production of
 RT fimbriectin- and Congo red-binding curli polymers in *Escherichia coli*
 RT K-12.";
 RL Mol. Microbiol. 18:661-670(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Blythe M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of *Escherichia coli* K-12.";
 RL Science 277:1232-1244(1997).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97061202; PubMed=8905232;
 RA Oshima T., Aiba T., Baba T., Fujita K., Hayashi K., Honjo A.,
 RA Ikemoto K., Inada T., Itoh T., Kajinaka M., Kanai K., Kashimoto K.,
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
 RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horuchi T.;
 RT "A 718-kb DNA sequence of the *Escherichia coli* K-12 genome
 RT corresponding to the 12.7-28.0 min region on the linkage map.";
 RL DNA Res. 3:137-155(1996).
 RN [5]
 RP SEQUENCE OF 21-40.
 RC STRAIN=K12 / YMEB;
 RX MEDLINE=93023873; PubMed=1357528;
 RA Argyris A., Olsen A., Pfeiffer J., Russell D.G., Normark S.;
 RT "The Cti protein activates cryptic genes for curli formation and
 RT fimbriectin binding in *Escherichia coli* HB101.";
 RL Mol. Microbiol. 6:2443-2452(1992).
 RN [6]
 RP SEQUENCE OF 21-31.
 RC MEDLINE=91310586; PubMed=1677357;
 RA Collinson S.K., Emeody L., Trust T.J., Kay W.W.;
 RT "Purification and characterization of thin, aggregative fimbriae from
 RT *Salmonella enteritidis*.";
 RL J. Bacteriol. 173:4773-4781(1991).
 CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 CC COLLED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 CC FIBRONECTIN.

```

CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L04979; AAA23616.1; -
DR EMBL; X90754; CAA52282.1; -
DR EMBL; AE000205; AAC74126.1; -
DR EMBL; D90741; BAA35832.1; -
DR EMBL; D90742; BAA35840.1; -
DR PIR; S70788; S70788.
DR Ecogene; EG11489; CSGA.
KM Fimbrtia; Signal; Complete proteome.
FT SIGNAL 1 20
FT CHAIN 21 151 MAJOR CURLIN SUBUNIT.
FT CONFLICT 7 7 A -> E (IN REF. 1).
SQ SEQUENCE 151 AA; 15049 MW; C003470D208D395F CRC64;

Query Match 68.8%; Score 533; DB 1; Length 151;
Best Local Similarity 70.2%; Pred. No. 2,4e-38;
Matches 106; Conservative 18; Mismatches 27; Indels 0; Gaps 0;

QY 1 MLLKVAAPAAIVSGSALAGVPPWGGGNNHGGNSGPDSTLSIYQGSANALAL 60
DB 1 MLLKVAAPAAIVSGSALAGVPPWGGGNNHGGNSGPDSTLSIYQGSANALAL 60
QY 61 SDARKSETTITQSGYNGADVGGADNSTIELTQNGFNNATIDQNAKNTDQVTRVVT 120
DB 61 TDARNSDLTITQHGCGNGADVGGSDSSIDLTRGFNSATLIDQNGKNSMTVQFGG 120

QY 121 HEMAHANQTSADSSVVRQVGFNNATNAY 151
DB 121 GNGAAYDQTASNSVTVTVQVFGNNATNAY 151

RESULT 3
CSGA_ECO57 STANDARD; PRT; 152 AA.
ID Q93U24;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Major curlin subunit precursor.
OS CSGA OR Z1676 OR ECS1420.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
NCBI_TaxID=83334;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / ATCC 43895;
RX MEDLINE=21218556; PubMed=11319125;
RA Uhlich G.A., Keen J.E., Elder R.O.;
RT "Mutations in the csgD promoter associated with variations in curli
RT expression in certain strains of Escherichia coli O157:H7."
RT Appl. Environ. Microbiol. 67:2367-2370(2001).
RL (2)
RN SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postel G., Hackett J., Klink A., Boutin A., Shao Y., Miller L.,
RA Apodaca E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
RT Nature 409:529-533(2001).

```

```

RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasaki C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12."
RL DNA Res. 8:11-22(2001).
CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN.
CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE275733; AAK53212.1; -
DR EMBL; AE005315; AAG55788.1; -
DR EMBL; AP002554; BAB34843.1; -
DR PIR; D90806; D90806.
DR PIR; H85665; H85665.
KM Fimbrtia; Signal; Complete proteome.
FT SIGNAL 1 20 BY SIMILARITY.
FT CHAIN 21 152 MAJOR CURLIN SUBUNIT.
SQ SEQUENCE 152 AA; 15099 MW; EE2D2D94DDE91243 CRC64;

Query Match 66.0%; Score 511.5; DB 1; Length 152;
Best Local Similarity 68.4%; Pred. No. 1.5e-36;
Matches 104; Conservative 19; Mismatches 28; Indels 1; Gaps 1;

QY 1 MLLKVAAPAAIVSGSALAGVPPWGGGNNHGGNSGPDSTLSIYQGSANALAL 59
DB 1 MLLKVAAPAAIVSGSALAGVPPWGGGNNHGGNSGPDSTLSIYQGSANALAL 60
QY 60 QSDARKSETTITQSGYNGADVGGADNSTIELTQNGFNNATIDQNAKNTDQVTRVVT 119
DB 61 QADARNSDLTITQHGCGNGADVGGSDSSIDLTRGFNSATLIDQNGKNSMTVQFGG 120

QY 120 THEMAHANQTSADSSVVRQVGFNNATNAY 151
DB 121 GNGAAYDQTASNSVTVTVQVFGNNATNAY 152

RESULT 4
CSGB_SALTI STANDARD; PRT; 151 AA.
ID Q8Z7M3;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Minor curlin subunit precursor.
OS CSGA OR STY1180 OR T1777.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
NCBI_TaxID=601;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parthill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Comerton P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,

```

RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
 RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gara P., Parry C.,
 RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
 RA Whitehead S., Barrrell B.G., Brown J., Clark R., Cowling R.,
 RT "Complete genome sequence of a multiple drug resistant *Salmonella*
 RT enterica serovar Typhimurium";
 RL Nature 413:848-852(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Ty2 / ATCC 700931;
 RX MEDLINE=2531367; PubMed=12644504;
 RA Deng W., Liou S.-R., Plummet G., III, Mayhew G.F., Rose D.J.,
 RA Briand V., Kodoyanni V., Schwartz D.C., Blatter F.R.,
 RT "Comparative genomics of *Salmonella enterica* serovar Typhimurium
 RT and CT18";
 RL J. Bacteriol. 185:2330-2337(2003).
 CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
 CC CURLIN MONOMERS.
 CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 DR EMBL; AL627269; CAD08267.1; -
 DR EMBL; AE016840; AA069400.1; -
 KW Fimbrin; Signal; Complete proteome.
 FT SIGNAL 1 21
 FT CHAIN 22 151 MINOR CURLIN SUBUNIT.
 FT SEQUENCE 151 AA; 16254 MW; 161C5432CE573495 CRC64;
 SQ
 Query Match 12.7%; Score 98.5; DB 1; Length 151;
 Best Local Similarity 28.8%; Pred. No. 0.099;
 Matches 34; Conservative 17; Mismatches 48; Indels 19; Gaps 5;
 QY 51 GSANAAALALOSDARKE-----TTITSGYNGGADVGO-GADNST-----TELQ 94
 DB 14 GAPGATATNTVDLARSEYFVAVNELSKSFNOAIIQGVGTDSARVROGSKLLSVTSQ 73
 QY 95 NGFRNATITDQWAKNTD-QLVTRVVTHEMAHNAQTASDSSVWVRQVGFNNATANY 151
 DB 74 EGNNAKAVDQ--AGNYNFAYIEQTGNANDASISQSAVNSAIIQSGSKNKANITQY 129
 RESULT 5
 CSGA_SALTY STANDARD; PRT; 151 AA.
 ID CSGA_SALTY STANDARD; PRT; 151 AA.
 AC P55226;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Minor curlin subunit precursor (Fimbrin SEF17 minor subunit).
 GN CSGA OR AGFB OR STM1143.
 OS *Salmonella typhimurium*, and
 OS *Salmonella enteritidis*.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; *Salmonella*.
 OX NCBI_TaxID=602, 592;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S.typhimurium; STRAIN=SR-11;
 RX MEDLINE=98117058; PubMed=9457880;
 RA Romling U., Bian Z., Hammar M., Sieralta W.D., Normark S.,
 RT "Curli fibers are highly conserved between *Salmonella typhimurium* and
 RT *Escherichia coli* with respect to operon structure and regulation.";
 RL J. Bacteriol. 180:722-731(1998).

RN [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S.typhimurium; STRAIN=LT2 / SSGC1412 / ATCC 700720;
 RX MEDLINE=21534948; PubMed=11677609;
 RA McClelland M., Sanderson K.R., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott C., Holmes A., Grewal N., Mulvaney E.,
 RA Ryan B., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RT "Complete genome sequence of *Salmonella enterica* serovar Typhimurium
 RT LT2";
 RL Nature 413:852-856(2001).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S.enteritidis; STRAIN=27655-3B;
 RX MEDLINE=96146512; PubMed=8550497;
 RA Collinson S.K., Clouthier S.C., Doran J.L., Banerjee P.A., Kay W.W.,
 RT "Salmonella enteritidis agfBAC operon encoding thin, aggregative
 RT fimbriae";
 RL J. Bacteriol. 178:662-667(1996).
 CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
 CC CURLIN MONOMERS.
 CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 DR EMBL; AJ002301; CA05316.1; -
 DR EMBL; AE008749; AAL20073.1; -
 DR EMBL; U43280; AAC43598.1; -
 DR PIR; JC6040; JC6040.
 DR StyGene; SG10609; csGB.
 KW Fimbrin; Signal; Complete proteome.
 FT SIGNAL 1 21
 FT CHAIN 22 151 MINOR CURLIN SUBUNIT.
 FT SEQUENCE 151 AA; 16182 MW; C0FC5430E6DD361D CRC64;
 SQ
 Query Match 12.7%; Score 98.5; DB 1; Length 151;
 Best Local Similarity 28.8%; Pred. No. 0.099;
 Matches 34; Conservative 17; Mismatches 48; Indels 19; Gaps 5;
 QY 51 GSANAAALALOSDARKE-----TTITSGYNGGADVGO-GADNST-----TELQ 94
 DB 14 GAPGATATNTVDLARSEYFVAVNELSKSFNOAIIQGVGTDSARVROGSKLLSVTSQ 73
 QY 95 NGFRNATITDQWAKNTD-QLVTRVVTHEMAHNAQTASDSSVWVRQVGFNNATANY 151
 DB 74 EGNNAKAVDQ--AGNYNFAYIEQTGNANDASISQSAVNSAIIQSGSKNKANITQY 129
 RESULT 6
 CSGA_ECOLI STANDARD; PRT; 151 AA.
 ID CSGA_ECOLI STANDARD; PRT; 151 AA.
 AC P39828;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Minor curlin subunit precursor.
 GN CSGA OR B1041 OR Z1675 OR ECSI419.
 OS *Escherichia coli*, and
 OS *Escherichia coli* O157:H7.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; *Escherichia*.
 OX NCBI_TaxID=562, 83334;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / WC4100;
 RX MEDLINE=9641468; PubMed=8817489;
 RA Hammar M., Arngvist A., Bian Z., Olsen A., Normark S.;
 RT "Expression of two csg operons is required for production of
 RT fibronectin- and congo red-binding curli polymers in *Escherichia coli*
 RT K-12.";
 RL Mol. Microbiol. 18:661-670(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1555;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G., III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of *Escherichia coli* K-12.";
 RL Science 277:1453-1474(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97061202; PubMed=8905232;
 RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
 RA Ikemura K., Inada T., Itch T., Kajihara M., Kanai K., Kashimuro K.,
 RA Kimura S., Kitegawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
 RA Samed G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horinouchi T.;
 RT "A 718-kb DNA sequence of the *Escherichia coli* K-12 genome
 RT corresponding to the 12.7-28.0 min region on the linkage map.";
 RL DNA Res. 3:1137-115(1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G., III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potomnis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.";
 RL Nature 409:529-533(2001).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,
 RA Kuhara S., Shiba T., Hattori M., Shingawa H.;
 RT "Complete genome sequence of enterohemorrhagic *Escherichia coli*
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22(2001).
 RN [6]
 RP SEQUENCE OF 1-21 FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=95157246; PubMed=7854117;
 RA Arngvist A., Olsen A., Normark S.;
 RT "Sigma S-dependent growth-phase induction of the *csgBA* promoter in
 RT *Escherichia coli* can be achieved in vivo by sigma 70 in the absence
 RT of the nucleoid-associated protein H-NS.";
 RL Mol. Microbiol. 13:1021-1032(1994).
 CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
 CC CURLIN MONOMERS.
 CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL, X80754; CAA62281.1; -
 DR EMBL, AE000205; AAC74125.1; -
 DR EMBL, D90741; BAA35831.1; -
 DR EMBL, AE005315; AAG55787.1; -
 DR EMBL, AF002554; BAB34842.1; -
 DR PIR, C90806; C90806.
 DR PIR, G85665; G85665.
 DR PIR, S70787; S70787.
 DR Ecogene; EG12621; CSGB.
 KW Fimbrin; Signal; Complete proteome.
 FT SIGNAL
 FT CHAIN 1 21
 FT CHAIN 1 21
 SQ SEQUENCE 151 AA; 15882 MW; B1BD266B964014B8 CRC64;
 Query Match 12.5%; Score 96.5; DB 1; Length 151;
 Best Local Similarity 25.6%; Pred. No. 0.15;
 Matches 30; Conservative 24; Mismatches 46; Indels 17; Gaps 4;
 QY 51 GSANMALAQSPARKSE-----TTTQSGYGGADVGCGADNSTIELTNGFRNATTIQ 105
 DB 14 GAPGIAAAGYDIANSEVNFVAVNELSKSFNOAATIGAGTNNASQLRGGSGLAVVAQ 73
 QY 106 WNAKYVDLVTVTHEMAHANQ--TASDSV-----NVRQVFGNNATANOY 151
 DB 74 EGSSNRK-IDQTGDNVLAITYIDAGSANDASISQAGYNTAMITIOKSGSKRANTIOY 129
 RESULT 7
 OVO DROME STANDARD; PRT; 1028 AA.
 ID OVO DROME
 AC P1521; O9XZU4;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE OVO protein (Shaven baby protein).
 GN OVO OR SVB.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Ovary;
 RX MEDLINE=95021209; PubMed=7935398;
 RA Garfinkel M.D., Wang J., Liang Y., Mahowald A.P.;
 RT "Multiple products from the shavenbaby-ovo gene region of *Drosophila*
 RT melanogaster: relationship to genetic complexity.";
 RL Mol. Cell. Biol. 14:6809-6818(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Oregon-R;
 RX MEDLINE=91293102; PubMed=1712294;
 RA Mevel-Ninio M.T.M., Terracol R., Kafatos F.C.;
 RT "The ovo gene of *Drosophila* encodes a zinc finger protein required
 RT for female germ line development.";
 RL EMBO J. 10:2259-2266(1991).
 CC -1- FUNCTION: REQUIRED FOR SURVIVAL AND DIFFERENTIATION OF FEMALE GERM
 CC LINE CELLS. PLAYS A ROLE IN GERM LINE SEX DETERMINATION.
 CC -1- SUBCELLULAR LOCATION: Nuclear (potential).
 CC -1- DEVELOPMENTAL STAGE: FIRST APPEARS IN THE GERMARIIUM AND
 CC ACCUMULATES IN NURSE CELLS DURING OOGENESIS. STORED IN THE EGG,
 CC BUT IS RAPIDLY LOST IN THE EMBRYOS EXCEPT FOR ITS CONTINUED
 CC PRESENCE IN THE GERM LINE PRECURSOR POLE CELLS.
 CC -1- SIMILARITY: Contains 4 C2H2-type zinc fingers.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; U11833; AAB60216.1; ALT_SEQ.
 DR EMBL; X59772; CAB36921.1; ALT_SEQ.
 DR PIR; A56038; A56038.
 DR HSSP; P07248; 2ADR.
 DR TRANSFAC; T00669; .
 DR FLYBASE; Fgn0003028; ovo.
 DR InterPro; IPR007087; Znf_C2H2.
 DR Pfam; PF00096; zf-C2H2; 3.
 DR SMART; SM00355; Znf_C2H2; 4.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
 DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 3.
 DR Zinc-finger; Metal-binding; DNA-binding; Repeat; Nuclear protein;
 KM Transcription regulation.
 FT DOMAIN 62 66 POLY-ALA.
 FT DOMAIN 72 77 POLY-GLY.
 FT DOMAIN 80 85 POLY-GLY.
 FT DOMAIN 98 108 POLY-GLY.
 FT DOMAIN 144 152 POLY-HIS.
 FT DOMAIN 153 159 POLY-ASN.
 FT DOMAIN 336 339 POLY-GLN.
 FT DOMAIN 347 353 POLY-GLN.
 FT DOMAIN 357 361 POLY-GLN.
 FT DOMAIN 410 414 POLY-GLN.
 FT DOMAIN 418 422 POLY-GLN.
 FT DOMAIN 426 432 POLY-GLN.
 FT DOMAIN 445 453 POLY-GLN.
 FT DOMAIN 456 459 POLY-GLN.
 FT DOMAIN 466 474 POLY-GLN.
 FT DOMAIN 497 517 POLY-ALA.
 FT DOMAIN 524 529 POLY-SER.
 FT DOMAIN 549 558 POLY-ALA.
 FT DOMAIN 639 651 POLY-ALA.
 FT DOMAIN 717 725 POLY-ALA.
 FT DOMAIN 797 802 POLY-GLN.
 FT DOMAIN 820 823 POLY-GLN.
 FT DOMAIN 826 832 POLY-GLN.
 FT ZN_FING 874 886 C2H2-TYPE 1.
 FT ZN_FING 902 924 C2H2-TYPE 2.
 FT ZN_FING 930 953 C2H2-TYPE 3.
 FT ZN_FING 969 992 C2H2-TYPE 4.
 FT CONFLICT 647 647 A -> R (IN REF. 2).
 SQ SEQUENCE 1028 AA; 110620 MW; D7068BB2BC0FF77 CRC64;

Query Match 11.9%; Score 92.5; DB 1; Length 1028;
 Best Local Similarity 26.7%; Pred. No. 2.8;
 Matches 40; Conservative 14; Mismatches 61; Indels 35; Gaps 6;

QY 3 LKVAAPALIVVSGSALAGVVPQWGGGNNHGGSSGPDSTLSTIYYGSANMAALAQSD 62
 Db LKVAAPALIVVSGSALAGVVPQWGGGNNHGGSSGPDSTLSTIYYGSANMAALAQSD 62
 59 LQNAAAAYMSASSG-----GGGCTGGGGGASGPGGPGPANSGGGGGG----- 104
 QY 63 ARKSEITITGSGYGAGDVGGAGDNSTIELTONGFRNNATIDQNAANYDQLVTRVYTHE 122
 Db ARKSEITITGSGYGAGDVGGAGDNSTIELTONGFRNNATIDQNAANYDQLVTRVYTHE 122
 105 -----GGNGYINCQGVG-GPNNNS---LDGNLNLAFASVSNYNEEN-----SKPHNH 147
 QY 123 MAHA-----NQYASSSVTVVQVQVGGNATA 148
 Db MAHA-----NQYASSSVTVVQVQVGGNATA 148
 148 HHQHNNNNNNNGQTMGHPFGGNPSA 177
 Db HHQHNNNNNNNGQTMGHPFGGNPSA 177

RESULT 8
 ID TNK1_HUMAN STANDARD; PRT; 1327 AA.
 AC 095271; 095272;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Tankyrase 1 (EC 2.4.2.30) (TNK1) (Tankyrase 1) (TNKS-1) (TRF1-
 DE interacting ankyrin-related ADP-ribose polymerase).
 GN TNKS OR TNKS1 OR TINI OR TINF1 OR PARP.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 CX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
 RC TISSUE=Testis;
 RX MEDLINE=99040105; PubMed=9822378;
 RA Smith S., Giriat L., Schmitt A., de Lange T.;
 RT "Tankyrase, a poly(ADP-ribose) polymerase at human telomeres.";
 RL Science 282:1484-1487(1998).
 RN [2]
 RP SUBCELLULAR LOCATION.
 RX MEDLINE=99454782; PubMed=10523501;
 RA Smith S., de Lange T.;
 RT "Cell cycle dependent localization of the telomeric PARP, tankyrase,
 RT to nuclear pore complexes and centrosomes.";
 RL J. Cell Sci. 112:3649-3656(1999).
 RN [3]
 RP FUNCTION, AND PHOSPHORYLATION.
 RX MEDLINE=20556282; PubMed=10988299;
 RA Chi N.-W., Lodish H.F.;
 RT "Tankyrase is a Golgi-associated mitogen-activated protein kinase
 RT substrate that interacts with IRAP in GLUT4 vesicles.";
 RL J. Biol. Chem. 275:38437-38444(2000).
 RN [4]
 RP FUNCTION, AND MUTAGENESIS OF HIS-1184 AND GLU-1291.
 RX MEDLINE=21602874; PubMed=11739745;
 RA Cook B.D., Dynek J.N., Chang W., Shostak G., Smith S.;
 RT "Role for the related poly(ADP-Ribose) polymerases tankyrase 1 and 2
 RT at human telomeres.";
 RL Mol. Cell. Biol. 22:332-342(2002).
 CC -1- FUNCTION: May regulate vesicle trafficking and modulate the
 CC subcellular distribution of SLC24/4/UT4-vesicles. Has PARP
 CC activity and can modify TRF1, and thereby contribute to the
 CC regulation of telomere length.
 CC -1- CATALYTIC ACTIVITY: NAD(+) + [ADP-D-riboseyl] (N)-acceptor =
 CC nicotinamide + (ADP-D-riboseyl) (N+1)-acceptor.
 CC -1- SUBUNIT: Oligomerizes and associates with TNKS2. Interacts with
 CC the cytoplasmic domain of LNP/PCase in SLC24/4/UT4-vesicles.
 CC Binds to the N-terminus of telomeric TRF1 via the ANK repeats.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic, associated with the Golgi and
 CC with junctional SLC24/4/UT4-vesicles. A minor proportion is
 CC also found at nuclear pore complexes and around the pericentriolar
 CC matrix of mitotic centrosomes. During interphase, a small fraction
 CC of TNKS is found in the nucleus, associated with TRF1.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC Name=2; IsoId=O95271-1; Sequence=Displayed;
 CC Name=O95271-2; Sequence=VSP_004538, VSP_004539;
 CC Note=No experimental confirmation available;
 CC -1- TISSUE SPECIFICITY: Ubiquitous; highest levels in testis.
 CC -1- PTM: Upon insulin-stimulation, phosphorylated on serine residues
 CC by MAPK kinases.
 CC -1- PTM: ADP-ribosylated (-auto).
 CC -1- SIMILARITY: Belongs to the PARP family.
 CC -1- SIMILARITY: Contains 15 ANK repeats.
 CC -1- SIMILARITY: Contains 1 sterile alpha motif (SAM) domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AF082556; AAC79841.1; -.


```

DR EMBL: AF082557; AAC79842.1; -
DR EMBL: AF082558; AAC79843.1; -
DR EMBL: AF082559; AAC79844.1; -
DR HSSP: Q00420; IAWC.
DR GENE: HGNC:11941; TNKS.
DR MIM: 603303; -
DR GO: GO:0000781; C:chromosome, telomeric region; IDA.
DR GO: GO:0000950; F:NAD ADP-ribosyltransferase activity; IDA.
DR GO: GO:000515; F:protein binding; IPI.
DR GO: GO:0007004; P:telomerase-dependent telomere maintenance; NAS.
DR InterPro: IPR002110; ANK.
DR InterPro: IPR001660; SAM.
DR Pfam: PF00023; ank; 19.
DR Pfam: PF00336; SAM; 1.
DR PRINTS: PR01415; ANKYRIN.
DR SMART: SM00248; ANK; 17.
DR SMART: SM00454; SAM; 1.
DR PROSITE: PSS0088; ANK_REPEAT; 15.
DR PROSITE: PSS0297; ANK_REPEAT_REGION; 1.
DR PROSITE: PSS0105; SAM_DOMAIN; 1.
DR Transferrase: Glycosyltransferase; NAD; Golgi stack; Telomere;
KW Nuclear protein; Repeat; ANK repeat; ADP-ribosylation;
KW Phosphorylation; Alternative splicing.
FT REPEAT 215 247 ANK 1.
FT REPEAT 248 280 ANK 2.
FT REPEAT 281 313 ANK 3.
FT REPEAT 368 400 ANK 4.
FT REPEAT 401 433 ANK 5.
FT REPEAT 434 466 ANK 6.
FT REPEAT 521 556 ANK 7.
FT REPEAT 557 589 ANK 8.
FT REPEAT 590 622 ANK 9.
FT REPEAT 683 715 ANK 10.
FT REPEAT 716 748 ANK 11.
FT REPEAT 749 781 ANK 12.
FT REPEAT 836 868 ANK 13.
FT REPEAT 869 901 ANK 14.
FT REPEAT 902 934 ANK 15.
FT DOMAIN 1030 1089 SAM.
FT DOMAIN 1176 1327 PARP.
FT DOMAIN 9 14 POLY-HIS.
FT DOMAIN 27 34 POLY-PRO.
FT DOMAIN 128 134 POLY-SER.
FT DOMAIN 137 145 POLY-SER.
FT VARSPLC 641 643 EST -> GHS (in isoform 2).
FT VARSPLC 644 1327 Missing (in isoform 2).
FT VARSPLC 644 1327 /FtId=VSP_004539.
FT MUTAGEN 1184 1184 /FtId=VSP_004539.
FT MUTAGEN 1291 1291 H->A: LOSS OF ACTIVITY; WHEN ASSOCIATED
WITH A-1291.
FT MUTAGEN 1291 1291 E->A: LOSS OF ACTIVITY; WHEN ASSOCIATED
WITH A-1184.
SQ SEQUENCE 1327 AA; 142010 MW; E14DE985C710B957 CRC64;

Query Match 11.7%; Score 90.5; DB 1; Length 1327;
Best Local Similarity 28.4%; Pred. No. 5.6;
Matches 42; Conservative 16; Mismatches 61; Indels 29; Gaps 6;

```

```

QY 6 VAAFAAI-VVSGSALAGVFPWGCGGNHNGGNSGPDSTLSIYOGSANAALALQSDAR 64
DB 99 VAAAPVPAVSTSSAAGVAPNPAGSGSNSSPSSPTSS-SSSPSSPSSSLAESPEEA 157
QY 65 KSEITIT-----OSGVNGADVGAGDNSTIELTONGFRNNATIDQWNAKVDLVTRVYT 120
DB 158 GVSSTNPPLGPAAGPRTGVPVNSGALRELEACRNG-----DVSRYKRLVDA--- 204
QY 121 HEMAHANQTASD-----SSVMVRYGVFG 143
DB 205 -----ANVAKDMAGRKSSPLHFAAGFG 227

RESULT 9
MSA2_PLAF2

```

```

ID MSA2_PLAF2 STANDARD; PRT; 347 AA.
AC Q03646;
DR 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Mezozoite surface antigen 2 precursor (MSA-2).
GN MSA2.
OS Plasmodium falciparum (isolate Nig32 / Nigeria).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_Taxid=70150;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91156685; PubMed=2000383;
RA Smythe J.A., Coppel R.L., Day K.P., Martin R.K., Oduola A.M.J.,
RA Kemp D.J., Anders R.P.;
RT "Structural diversity in the Plasmodium falciparum mezozoite surface
RT antigen 2."
RL Proc. Natl. Acad. Sci. U.S.A. 88:1751-1755(1991).
CC -1- FUNCTION: May play a role in the mezozoite attachment to the
CC erythrocyte.
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC (potential).
CC -1- DEVELOPMENTAL STAGE: During the trophozoite and schizont stages.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL: M59765; AAA29691.1; -.
DR PIR: B39112; B39112.
DR InterPro: IPR001136; MSA_2.
DR Pfam: PF00385; MSR_2; 1.
KW Malaria; Membrane; Glycoprotein; Antigen; Signal; Repeat;
KW GPI-anchor; Mezozoite.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 323 MEZOZOITE SURFACE ANTIGEN 2.
FT PROPEP 324 347 HYDROPHOBIC. REMOVED DURING MATURATION
(BY SIMILARITY).
FT DOMAIN 44 273 POLYMORPHIC REGION.
FT DOMAIN 177 184 POLY-THR.
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 36 36 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 224 224 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 236 236 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 320 320 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 321 321 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 347 AA; 33786 MW; ABCF24BBS60BF537 CRC64;

Query Match 11.6%; Score 90; DB 1; Length 347;
Best Local Similarity 24.3%; Pred. No. 1.3;
Matches 37; Conservative 16; Mismatches 71; Indels 28; Gaps 5;

```

```

QY 7 AAFATVVSAGLAGVFPWGCGGNHNGGNSGPDSTLSIYOGSANAALALQSDARKS 66
DB 74 AAGNAGVASAGNAGAGNAGAGNAGAGNAGAG-----NGAGNAG-----GNGAGNG 121
QY 67 ETTITSGYNGA--DVGGADNSTIELTONGFRNNATIDQWNAKVDQ-----LVTR 117
DB 122 AAGNAGAGNAGAGNAGAGNAGAGNAGAGNAGAGNAGAGNAGAGNAGAGNAGAGNAG 181
QY 118 VTHMAHANQTASDSSVMVRYGVFGNATAN 149
DB 182 TTTNDAASTSTSSNS-----NNNAETN 206

RESULT 10
VG38_BPT2 STANDARD; PRT; 262 AA.
ID VG38_BPT2
AC P07875;

```

```

DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 01-AUG-1990 (Rel. 15, Last annotation update)
DE Receptor recognizing protein (Protein Gp38) .
GN 38.
OS Bacteriophage T2.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
OC T4-like viruses.
OC NCBI_TaxId=10664;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87283911; PubMed=3302276;
RA Riede I., Drexler K., Eschbach M.L., Henning U.;
RA "DNA sequence of genes 38 encoding a receptor-recognizing protein of
RT bacteriophages T2, K3 and of K3 host range mutants.";
RT J. Mol. Biol. 194:31-39(1987)
CC -!- FUNCTION: Vg38 is at the tip of the long tail fibers and serves as
CC the phage recognition site for the cellular receptor.
CC -!- MISCELLANEOUS: THIS PHAGE USE OUTER MEMBRANE PROTEINS OMPF AND TTR
CC AS RECEPTORS.
CC -----
CC This SMS-S-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X05312; CA28935.1; -
DR PIR; S00275; S00275.
DR InterPro; IPR007932; Tail_fibre_GP38.
DR Pfam; PF05268; GP38; 1.
DR KMW fiber protein; Phage recognition.
SQ SEQUENCE 262 AA; 25801 MW; 0567366918F6C745 CRC64;

Query Match: 11.5%; Score 89; DB 1; Length 262;
Best Local Similarity 34.4%; Pred. NO. 1.2;
Matches. 32; Conservative 9; Mismatches 38; Indels 14; Gaps 5;

QY 27 GGGGNNHNGGNSGPDSTLSTLYQYSSANAALALOSDARKSETTTTQSGYGNGAGVGGACD 86
DB 175 GCGGPFVGVGKIGSDSLT-----GSNASL---TDACTGTTTF-QYGAANGGVAGCGG 225
QY 87 ---NSTIELTQNGPRNNAATIDQANAKATDQVLT 116
DB 226 RGMGRNVTYSEGAAGAATVG--NAPMNQNVGT 256

RESULT 11
GP63_LEIDO
ID _GP63_LEIDO STANDARD; PRT; 590 AA.
DC P23223;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Leishmanolysin precursor (BC 3.4.24.36) (Cell surface protease)
DE (Major surface glycoprotein) (Gp63 protein) (Promastigote surface
DE endopeptidase).
DE GP63.
OS Leishmania donovani.
OC Eukaryota; Eukleozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OC NCBI_TaxId=5661;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=LV9;
RX MEDLINE=92107220; PubMed=1762629;
RA Webb J.R., Button L.L., McMaster R.W.;
RA "Heterogeneity of the genes encoding the major surface glycoprotein
RT of Leishmania donovani.";
RT Mol. Biochem. Parasitol. 48:173-184(1991).
CC -!- FUNCTION: Has an integral role during the infection of macrophages
CC in the mammalian host.

```

```

CC -1- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and
CC P1', and basic residues at P2 and P3'. A model nonapeptide is
CC cleaved at -Ala-Tyr-|-Leu-Lys-Lys-.
CC CORRECTOR: Binds 1 zinc ion per subunit (By similarity).
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -1- SIMILARITY: Belongs to peptidase family M8.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC with the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL: M60048; AAA29244.1; -.
DR HSSP: P08148; 1MLT.
DR MEROPS: M08.001; -.
DR InterPro: IPR006025; Pept_M_Zn_BS.
DR InterPro: IPR001577; Peptidase_M8.
DR Pfam: PF01457; Peptidase_M8; 1.
DR PRINTS: PR00782; LSHMANOLYSIN.
DR PROSITE: PS00142; ZINC_PROTEASE; 1.
KM Hydroxylase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;
KW Zymogen; Signal; Cell adhesion; GPI-anchor; Lipoprotein.
FT SIGNAL 1 39
FT PROPEP 40 87
FT CHAIN 88 565
FT PROPEP 566 590
FT METAL 251 251
FT ACT_SITE 252 252
FT METAL 252 252
FT METAL 255 255
FT METAL 321 321
FT DISULFID 112 129
FT DISULFID 178 217
FT DISULFID 301 373
FT DISULFID 380 443
FT DISULFID 393 412
FT DISULFID 402 477
FT DISULFID 454 498
FT DISULFID 503 553
FT DISULFID 523 546
FT CARBOHYD 287 287
FT LIPID 565 565
SQ SEQUENCE 590 AA; 62950 MW; 0FB315D29659F58 CRC64;
Query Match 11.2%; Score 87; DB 1; Length 590;
Best Local Similarity 89.5%; Pred. No. 4.4;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 108 AKAYDQLVTRVVTHEMAHA 126
Db 238 ASRYDQLVTRVVTHEMAHA 256
RESULT 12
GP63_LEICH STANDARD; PRT; 599 AA.
AC P15706;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Leishmanolysin precursor (EC 3.4.24.36) (Cell surface protease)
DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface
DE endopeptidase).
GN GP63.
OS Leishmania chagasi.
OS Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OC NCBI_TaxID=44271;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90205976; PubMed=2320059;

```

RA Miller R.A., Reed S.G., Parsons M.;
 RT "Leishmania gp63 molecule implicated in cellular adhesion lacks an
 RT Arg-Gly-Asp sequence.";
 RL Mol. Biochem. Parasitol. 39:267-274 (1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=92112918; PubMed=1370484;
 RA Ramamoorthy R., Donelson J.E., Paetz K.E., Maybodi M., Roberts S.C.,
 RA Wilson M.E.;
 RT "Three distinct RNAs for the surface protease gp63 are differentially
 RT expressed during development of Leishmania donovani chagasi
 RT promastigotes to an infectious form.";
 RL J. Biol. Chem. 267:1888-1895 (1992).
 CC -1- FUNCTION: Has an integral role during the infection of macrophages
 CC in the mammalian host.
 CC -1- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and
 CC P1' and basic residues at P2 and P3'. A model nonapeptide is
 CC cleaved at -Ala-Tyr|-Leu-Lys-Lys-.
 CC -1- CORFACTOR: Binds 1 zinc ion per subunit (by similarity).
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
 CC -1- SIMILARITY: Belongs to peptidase family M8.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC DR EMBL; M80672; AAA29238.1; -;
 CC DR EMBL; M28527; AAA29235.1; -;
 CC DR PIR; A44951; A44951.
 CC DR HSSP; P08148; 1LMU.
 CC DR MEROPS; M08.001; -;
 CC DR InterPro; IPR006025; Pept_M_Zn_BS.
 CC DR InterPro; IPR001577; Peptidase_M8.
 CC DR Pfam; PF01457; Peptidase_M8; 1.
 CC DR PRINTS; PR00782; LSHMANOLYSIN.
 CC DR PROSITE; PS00142; ZINC_PROTEASE; 1.
 CC KM Hydroxylase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;
 CC Zymogen; Signal; Cell adhesion; GPI-anchor; Lipoprotein.
 CC KX SIGNAL 1 39
 CC FT PROPEP 40 97
 CC FT CHAIN 98 574
 CC FT PROPEP 575 599
 CC FT METAL 261 261
 CC FT ACT_SITE 262 262
 CC FT METAL 265 265
 CC FT METAL 331 331
 CC FT DISULFID 122 139
 CC FT DISULFID 188 227
 CC FT DISULFID 311 383
 CC FT DISULFID 390 452
 CC FT DISULFID 403 422
 CC FT DISULFID 412 486
 CC FT DISULFID 463 507
 CC FT DISULFID 512 562
 CC FT DISULFID 532 555
 CC FT CARBOHYD 297 297
 CC FT CARBOHYD 394 394
 CC FT LIPID 574 574
 CC SQ SEQUENCE 599 AA; 63848 MW; 746730A8E82A287C CRC64;
 CC Query Match 11.2%; Score 87; DB 1; Length 599;
 CC Best Local Similarity 89.5%; Pred. No. 4.4;
 CC Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

RESULT 13
 GP63 LEIMA
 ID GP63 LEIMA STANDARD; PRT; 602 AA.
 AC P08148; P15906;
 DT 01-ANG-1988 (Rel. 08, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Leishmanolysin precursor (EC 3.4.24.36) (Cell surface protease)
 DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface
 DE endopeptidase).
 GN GP63.
 OS Leishmania major.
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 OX NCBI_TaxID=5664;
 RX [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 101-123.
 RX MEDLINE=88154764; PubMed=3346625;
 RA Button L.L., McMaster W.R.;
 RT "Molecular cloning of the major surface antigen of Leishmania.";
 RL J. Exp. Med. 167:724-729 (1988).
 RN [2]
 RP REVISIONS.
 RA Button L.L., McMaster W.R.;
 RL J. Exp. Med. 171:589-589 (1990).
 RN [3]
 RP GPI-ANCHOR.
 RX MEDLINE=91009116; PubMed=2145267;
 RA Schneider P., Ferguson M.A.J., McConville M.J., Mehler A.,
 RA Homans S.W., Bordier C.;
 RT "Structure of the glycosyl-phosphatidylinositol membrane anchor of
 RT the Leishmania major promastigote surface protease.";
 RL J. Biol. Chem. 265:16955-16964 (1990).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).
 RX MEDLINE=95406217; PubMed=7675788;
 RA Schlagenhaut E., Etges R., Metcalf P.;
 RT "Crystallization and preliminary X-ray diffraction studies of
 RT Leishmanolysin, the major surface metalloprotease from Leishmania
 RT major.";
 RL Proteins 22:58-66 (1995).
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY (1.86 ANGSTROMS).
 RX MEDLINE=98416698; PubMed=9739094;
 RA Schlagenhaut E., Etges R., Metcalf P.;
 RT "The crystal structure of the Leishmania major surface proteinase
 RT Leishmanolysin.";
 RL Structure 6:1035-1046 (1998).
 CC -1- FUNCTION: Has an integral role during the infection of macrophages
 CC in the mammalian host.
 CC -1- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and
 CC P1' and basic residues at P2 and P3'. A model nonapeptide is
 CC cleaved at -Ala-Tyr|-Leu-Lys-Lys-.
 CC -1- CORFACTOR: Binds 1 zinc ion per subunit.
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
 CC -1- PTM: THE PHOSPHATIDYLINOSITOL MOIETY OF THE GPI-ANCHOR CONTAINS A
 CC FULLY SATURATED, UNBRANCHED 1-O-ALKYL CHAIN (MAINLY C24:0) AND A
 CC MIXTURE OF FULLY SATURATED UNBRANCHED 2-O-ACYL CHAINS (C12:0,
 CC C14:0, C16:0, AND C18:0).
 CC -1- SIMILARITY: Belongs to peptidase family M8.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC DR EMBL; Y00647; CAA68673.1; -;
 CC DR PIR; P10221; P10221.
 CC DR FDB; 1LMU; 17-SEP-97.
 CC DR MEROPS; M08.001; -;

DR InterPro; IPR006025; Pept M Zn BS;
 DR InterPro; IPR001577; Peptidase_M8;
 DR Pfam; PF01457; Peptidase_M8; 1.
 DR PRINTS; PR00782; LSHMANOLYSIN
 DR PROSITE; PS00142; ZINC_PROTEASE; 1.
 KW Hydrolyase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;
 KW Zymogen; Signal; Cell adhesion; GPI-anchor; 3D-structure; Lipoprotein.
 FT SIGNAL 1 39
 FT PROPEP 40 100
 FT CHAIN 101 577
 FT PROPEP 578 602
 FT METAL 264 264
 FT ACT_SITE 265 265
 FT METAL 268 268
 FT METAL 334 334
 FT DISULFID 125 142
 FT DISULFID 191 230
 FT DISULFID 314 386
 FT DISULFID 393 455
 FT DISULFID 406 425
 FT DISULFID 415 489
 FT DISULFID 466 510
 FT DISULFID 515 565
 FT DISULFID 535 558
 FT CARBOHYD 300 300
 FT CARBOHYD 407 407
 FT LIPID 577 577
 FT STRAND 101 102
 FT STRAND 107 108
 FT STRAND 111 114
 FT HELIX 116 119
 FT TURN 121 122
 FT TURN 128 129
 FT STRAND 131 133
 FT STRAND 139 141
 FT HELIX 144 146
 FT HELIX 150 158
 FT TURN 159 159
 FT HELIX 160 169
 FT TURN 170 171
 FT STRAND 172 174
 FT STRAND 177 178
 FT STRAND 180 181
 FT TURN 189 190
 FT HELIX 191 193
 FT HELIX 198 202
 FT TURN 203 203
 FT STRAND 205 206
 FT STRAND 210 215
 FT TURN 221 222
 FT STRAND 226 232
 FT TURN 234 235
 FT STRAND 238 244
 FT HELIX 247 249
 FT HELIX 256 269
 FT TURN 270 271
 FT HELIX 274 279
 FT TURN 280 281
 FT STRAND 283 286
 FT STRAND 289 291
 FT HELIX 296 299
 FT HELIX 302 312
 FT TURN 313 313
 FT TURN 315 316
 FT STRAND 320 322
 FT TURN 328 332
 FT STRAND 335 335
 FT TURN 337 339
 FT TURN 341 342
 FT STRAND 343 343
 FT TURN 344 345
 FT STRAND 353 353
 FT HELIX 356 364

POTENTIAL.
 ACTIVATION PEPTIDE.
 LEISHMANOLYSIN.
 REMOVED IN MATURE FORM.
 ZINC (CATALYTIC).
 ZINC (CATALYTIC).
 ZINC (CATALYTIC).
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 GPI-anchor amidated asparagine.

FT TURN 365 366
 FT STRAND 369 370
 FT HELIX 372 374
 FT TURN 380 383
 FT HELIX 386 390
 FT STRAND 394 395
 FT TURN 396 397
 FT STRAND 398 399
 FT TURN 402 404
 FT TURN 413 414
 FT TURN 417 418
 FT STRAND 421 425
 FT STRAND 428 429
 FT HELIX 435 437
 FT HELIX 443 444
 FT STRAND 445 446
 FT TURN 450 454
 FT STRAND 458 465
 FT TURN 466 467
 FT HELIX 470 472
 FT TURN 475 477
 FT HELIX 478 480
 FT TURN 485 486
 FT STRAND 487 494
 FT STRAND 496 496
 FT STRAND 506 516
 FT TURN 517 520
 FT STRAND 521 525
 FT TURN 527 528
 FT STRAND 533 534
 FT TURN 537 538
 FT STRAND 540 542
 FT HELIX 543 545
 FT TURN 546 546
 FT STRAND 550 550
 FT TURN 552 553
 FT STRAND 555 557
 FT HELIX 561 565
 FT TURN 566 567
 FT HELIX 569 572
 FT TURN 573 573

SEQUENCE 602 AA; 982EF3245D87C43E CRC64;
 MW: 63953
 Query Match 11.2%; Score 87; DB 1; Length 602;
 Best Local Similarity 89.5%; Pred. No. 4.5;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 108 AKNYDQVTRVVTHEMAHA 126
 Db 251 ASRYDQVTRVVTHEMAHA 269

RESULT 14
 GP63_LEIME
 ID GP63 LEIME STANDARD; PRT; 646 AA.
 AC P3150;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Leishmanolysin C1 precursor (EC 3.4.24.36) (Cell surface protease)
 DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface endopeptidase).
 GN GP63-C1.
 OS Leishmania mexicana.
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 OX NCBI_TaxID=5665;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MNYC/BZ/62/M379;
 RX MEDLINE=93149206; PubMed=8426614;
 RA Medina-Acosta E., Kares R.B., Russell D.G.;
 RT "Structurally distinct genes for the surface protease of Leishmania mexicana are developmentally regulated.";

RL Mol. Biochem. Parasitol. 57:31-46(1993).
 CC -1- FUNCTION: Has an integral role during the infection of macrophages
 CC in the mammalian host.
 CC -1- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and
 CC P1', and basic residues at P2 and P3'. A model nonapeptide is
 CC cleaved at -Ala-Tyr-Lys-Lys-.
 CC -1- CATALYTIC ACTIVITY: Cleavage of a zinc ion per subunit (By similarity).
 CC -1- DEVELOPMENTAL STAGE: Expressed in both the promastigote and the
 CC amastigote forms.
 CC -1- SIMILARITY: Belongs to peptidase family M8.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; X64394; CAA45733.1; -.
 CC DR PIR; S19916; S19916.
 CC DR HSP; P08148; 11ML.
 CC DR MEROPS; M08.001; -.
 CC DR Glycosylated; P43150; -.
 CC DR InterPro; IPR006025; Pept_M_Zn_BS.
 CC DR InterPro; IPR001577; Peptidase_M8.
 CC DR Pfam; PF01457; Peptidase_M8; 1.
 CC DR PRINTS; PR00782; LSHMANOLYSIN.
 CC DR PROSITE; PS00142; ZINC_PROTEASE; 1.
 CC KM Hydrolyase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;
 CC Zymogen; Signal; Cell adhesion; Multigene family.
 CC FT SIGNAL 1 39
 CC FT PROPEP 40 102
 CC FT CHAIN 103 646
 CC FT METAL 266 266
 CC FT ACT_SITE 267 267
 CC FT METAL 270 270
 CC FT METAL 336 336
 CC FT DISULFID 127 144
 CC FT DISULFID 193 232
 CC FT DISULFID 316 388
 CC FT DISULFID 395 458
 CC FT DISULFID 408 427
 CC FT DISULFID 417 492
 CC FT DISULFID 469 513
 CC FT DISULFID 518 568
 CC FT DISULFID 538 561
 CC FT CARBOHYD 86 86
 CC FT CARBOHYD 297 297
 CC FT CARBOHYD 399 399
 CC FT CARBOHYD 409 409
 CC FT CARBOHYD 433 433
 CC FT CARBOHYD 445 445
 CC FT CARBOHYD 466 466
 CC FT CARBOHYD 501 501
 CC FT CARBOHYD 501 501
 CC SQ SEQUENCE 646 AA; 69054 MW; FE448DDC78C10B0A CRC64;
 QY Query Match 11.2%; Score 87; DB 1; Length 646;
 QY Best Local Similarity 89.5%; Pred. No. 4.8;
 QY Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 108 AKNYDQLVTVVTHENMAHA 126
 DB 253 ASRYDQLVTVVTHENMAHA 271

RESULT 15
 OMPB_RICJA STANDARD; PRT; 1656 AA.
 ID OMPB_RICJA
 AC 006653;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Outer membrane protein B precursor (168 kDa surface-layer protein)
 DE (Surface protein antigen) (Cell surface antigen 5) (Scs5) (rOmpB)
 DE (Omp B) [contains: 120 kDa surface-exposed protein (Surface protein
 DE antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide].
 GN OMPB
 OS Rickettsia japonica.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 OX NCBI_TaxID=35790;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TH;
 RA Uchiyama T.;
 RT "Sequencing of the gene encoding the protein rOmp B of Rickettsia
 RT japonica.";
 RL Submitted (MAY-1997) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: THE 120 kDa SURFACE-EXPOSED PROTEIN IS A MAJOR
 CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
 CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (By
 CC similarity).
 CC -1- FUNCTION: THE 32 kDa BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR
 CC (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-
 CC layer with hexagonal symmetry.
 CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; AB003681; BAA20138.1; -.
 CC DR InterPro; IPR006315; Autotransporter.
 CC DR InterPro; IPR005546; Autotransporter.
 CC DR Pfam; PF03797; Autotransporter; 1.
 CC DR TIGRFAMs; TIGR01414; autotrans_bar1; 2.
 CC KM Antigen; S-layer; Cell wall. 120 kDa SURFACE-EXPOSED PROTEIN.
 CC FT CHAIN 1 1338
 CC FT CHAIN 1339 1656
 CC FT DOMAIN 528 533
 CC FT POLY-GLY.
 CC SQ SEQUENCE 1656 AA; 168097 MW; 3132A69C9DD5999F CRC64;

QY Query Match 11.2%; Score 87; DB 1; Length 1656;
 QY Best Local Similarity 29.6%; Pred. No. 14;
 QY Matches 45; Conservative 14; Mismatches 55; Indels 38; Gaps 8;
 QY 6 VAAPAIIVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGANALALQSDARK 65
 DB 509 VLAAGATILDSATL-----TGDIQNGGG-----GAALQSTILANDARK 547
 QY 66 SETTITQSG---YNGADVGGAGDNSTIELTQNGFRNNATITDQNNAKKYDQLVTVVTH 121
 DB 548 ---TTLGGANIISANGTINFGANGTIRKIST--QNNIVVDCDLAIATDQ--TGVDVA 600
 QY 122 EWAHANQASDSVWVRQGF--GNMATTANOV 151
 DB 601 SSLTVAQTUTLTISGT---IGIIGANNVTLDQGF 628

Search completed: March 11, 2004, 18:34:58
 Job time : 7.3 secs

THIS PAGE BLANK (USPTO)


```

Db      121 NNAALVNGTASDSSVWVROVGFGNNAATANQY 151
      : |||||
RESULT 2
Q7X243      PRELIMINARY;      PRT;      150 AA.
AC      Q7X243;
DT      01-OCT-2003 (TREMBlrel. 25, Created)
DT      01-OCT-2003 (TREMBlrel. 25, Last sequence update)
DT      01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE      Curliin-csgA protein.
GN      CSGA.
OS      Citrobacter sp. Fec2.
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OX      Enterobacteriaceae; Citrobacter.
RN      [1]
RP      STRAIN=Fec2;
RC      SEQUENCE FROM N.A.
RA      Zogaj X., Bokranz W., Nimtz M., Romling U.;
RT      "Production of Cellulose and Curli Fimbriae by Members of the Family
RL      Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
DR      EMBL; AJ515700; CAD56672.1; -.
SQ      SEQUENCE 150 AA; 15016 MW; 1D7141BD6973DC6 CRC64;

Query Match      76.3%; Score 591.5; DB 2; Length 150;
Best Local Similarity 78.8%; Pred. No. 2.3e-39;
Matches 119; Conservative 12; Mismatches 19; Indels 1; Gaps 1;

Qy      1 MLLKVAFAAIVVSGSALAGVVPQWGGGNNHGGSGSDSTLSIYOGSANAALALQ 60
      : |||||
Db      1 MLLKVAFAAIVVSGSALAGVVPQWGGGNNHGGSGSDSTLSIYOGSANAALALQ 59
      : |||||
Qy      61 SDARKSETTITGSGYGNAGADVGGADNSTIELTONGFRNNAITDQNAKYDQLVTRVVT 120
      : |||||
Db      60 SDARKSDTTIHQNGGNGADVGGSDNSTIDLTONGFRKNATTIDQNGKSDITVSCYGG 119
      : |||||
Qy      121 HEMAANOTASDSSVWVROVGFGNNAATANQY 151
      : |||||
Db      120 HNAALVNGTASDSSVWVROVGFGNNAATANQY 150
      : |||||

RESULT 3
Q7X240      PRELIMINARY;      PRT;      149 AA.
AC      Q7X240;
DT      01-OCT-2003 (TREMBlrel. 25, Created)
DT      01-OCT-2003 (TREMBlrel. 25, Last sequence update)
DT      01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE      Curliin-csgA protein.
GN      CSGA.
OS      Citrobacter freundii.
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OX      Enterobacteriaceae; Citrobacter.
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=Fec4;
RA      Zogaj X., Bokranz W., Nimtz M., Romling U.;
RT      "Production of Cellulose and Curli Fimbriae by Members of the Family
RL      Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
DR      EMBL; AJ515701; CAD56675.1; -.
SQ      SEQUENCE 149 AA; 15260 MW; 946DD52017F648FD CRC64;

Query Match      71.4%; Score 553; DB 2; Length 149;
Best Local Similarity 74.8%; Pred. No. 2.5e-36;
Matches 113; Conservative 15; Mismatches 21; Indels 2; Gaps 1;

Qy      1 MLLKVAFAAIVVSGSALAGVVPQWGGGNNHGGSGSDSTLSIYOGSANAALALQ 60
      : |||||

```

```

Db      1 MLLKVAFAAIVVSGSALAGVVPQW--GGNNHGGSGNYGPDSLSIYOGSNNANAALQ 58
      : |||||
Qy      61 SDARKSETTITGSGYGNAGADVGGADNSTIELTONGFRNNAITDQNAKYDQLVTRVVT 120
      : |||||
Db      59 SDARKSDVITTHQGRNGAVVGGADNSTLSKOTGFONSATIDQNAKADISVTFQFG 118
      : |||||
Qy      121 HEMAANOTASDSSVWVROVGFGNNAATANQY 151
      : |||||
Db      119 RNGALVNGTASDSSVWVROVGFGNNAATANQY 149
      : |||||

RESULT 4
Q8CW63
AC      Q8CW63;
DT      01-MAR-2003 (TREMBlrel. 23, Created)
DT      01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT      01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE      Major curliin subunit precursor.
GN      CSGA OR C1306.
OS      Escherichia coli O6.
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OX      Enterobacteriaceae; Escherichia.
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=O6:HI / ATCC 700928;
RX      MEDLINE=22388234; PubMed=12471157;
RA      Welch R.A., Burland V., Plunkett G., III, Redford P., Roesch P.,
RA      Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA      Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA      Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT      "Extensive mosaic structure revealed by the complete genome sequence
RL      of uropathogenic Escherichia coli.";
DR      Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).
DR      EMBL; AE016759; AAN79779.1; -.
KW      Complete proteome.
SQ      SEQUENCE 152 AA; 15064 MW; 3BA57F34C1240E83 CRC64;

Query Match      65.6%; Score 508.5; DB 16; Length 152;
Best Local Similarity 68.4%; Pred. No. 8.5e-33;
Matches 104; Conservative 18; Mismatches 23; Indels 1; Gaps 1;

Qy      1 MLLKVAFAAIVVSGSALAGVVPQW--GGGNNHGGSGSGSDSTLSIYOGSANAALALQ 59
      : |||||
Db      1 MLLKVAFAAIVVSGSALAGVVPQYGGGNGHGGGNNHGGSGNSSELNITYOGGNSALALQ 60
      : |||||
Qy      60 GSDARKSETTITGSGYGNAGADVGGADNSTIELTONGFRNNAITDQNAKYDQLVTRVVT 119
      : |||||
Db      61 QADARNSDTTIHQGGGNGADVGGSDNSTIDLTORFGNSATIDQNGKSDITVTKQFG 120
      : |||||
Qy      120 THEMAMNOTASDSSVWVROVGFGNNAATANQY 151
      : |||||
Db      121 GGNGAADVOTASDSSVWVROVGFGNNAATANQY 152
      : |||||

RESULT 5
Q7X237
AC      Q7X237;
DT      01-OCT-2003 (TREMBlrel. 25, Created)
DT      01-OCT-2003 (TREMBlrel. 25, Last sequence update)
DT      01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE      Curliin-csgA protein.
GN      CSGA.
OS      Enterobacter sakazakii.
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OX      Enterobacteriaceae; Enterobacter.
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=Fec39;
RA      Zogaj X., Bokranz W., Nimtz M., Romling U.;

```



```

RT "Production of Cellulose and Curli Fimbriae by Members of the Family
  Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158(2003).
DR EMBL; AJ515702; CAD56678.1; -.
SQ SEQUENCE 150 AA; 15112 MW; 5D8B2D872DF1SF3 CRC64;

Query Match
  Best Local Similarity 55.7%; Score 431.5; DB 2; Length 150;
  Matches 90; Conservative 27; Mismatches 33; Indels 1; Gaps 1;

QY 1 MLLKVAAPAAIVSGSALAGVVPWGCGGNGHGGGSGPDSLTIYOGSANAALALQ 60
  DB 1 MRFKVAALAAIVSSSANAAGMTNQ-GGNGHGHGCGYGGPNSLTINLYOGGNSALALQ 59
QY 61 SPARKSETTITGSGYGADVGAGADNSTIELTONGFRNNATIDOWNAKNYDQVTRVVT 120
  DB 60 TDBRNSVLINISQGGGNGADVGOGSDSSINLTONGGNSATIDOWNSKDSVNVVGG 119
QY 121 HEAAHANQTRASDSSVWVRQVGFNNATANDY 151
  DB 120 LNALVDTASNSTVAVTQIGFNHATAHQY 150

RESULT 6
Q54069 PRELIMINARY; PRT; 76 AA.
ID 054069;
AC 054069;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE SEF17 fimbria (fragment).
GN AGFA.
OS Salmonella enteritidis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
  Enterobacteriaceae; Salmonella.
NCBI_TaxId=592;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=SE30;
RA Cox J.M., Eagles S., Woolcock J.B.;
RT "virulence of Salmonella enteritidis in chickens correlates with
  colony morphology and expression of SEF17 fimbriae.";
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U53207; AAA98671.1; -.
FT NON TER 1 1
FT NON TER 76 76
SQ SEQUENCE 76 AA; 7704 MW; 2FD5411241A7CB1 CRC64;

Query Match
  Best Local Similarity 49.7%; Score 385; DB 2; Length 76;
  Matches 75; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 30 GNNNGGNSGSPSTSIYOGSANAALALQSPARKSETTITGSGYGADVGAGADNST 89
  DB 1 GNNXGGGNSGSPSTSIYOGSANAALALQSPARKSETTITGSGYGADVGAGADNST 60
QY 90 IELTONGFRNNATIDQ 105
  DB 61 IELTONGFRNNATIDQ 76

RESULT 7
Q953J5 PRELIMINARY; PRT; 29 AA.
ID 0953J5;
AC 0953J5;
DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (T-EMBLrel. 16, Last annotation update)
DE Curli subunit monomer (fragment).
GN CSGA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
  Enterobacteriaceae; Escherichia.

```

```

OX NCBI_TaxId=562;
RN (1)
RP SEQUENCE FROM N.A.
RC TRANSPOSON-insertion sequence ISI;
RX MEDLINE=99314153; PubMed=10386375;
RA Ragione R.M., Collighan R.J., Woodward M.J.;
RT "Non-curliation of Escherichia coli O78:K80 isolates associated with
  ISI inserti on in cs9 and reduced persistence in poultry infection.";
RL FEMS Microbiol. Lett. 175:247-253(1999).
DR EMBL; AJ131756; CAB45380.1; -.
FT NON TER 29 29
SQ SEQUENCE 29 AA; 2789 MW; E290DFC07ABBB243 CRC64;

Query Match
  Best Local Similarity 15.7%; Score 122; DB 2; Length 29;
  Matches 26; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MLLKVAAPAAIVSGSALAGVVPWGCGG 29
  DB 1 MLLKVAALAAIVSSSALAGVVPYGGG 29

RESULT 8
Q8EIH4 PRELIMINARY; PRT; 502 AA.
ID 08EIH4;
AC 08EIH4;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Conserved hypothetical protein.
GN S00865.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
  Alteromonadaceae; Shewanella.
NCBI_TaxId=70863;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=MR-1;
RX MEDLINE=22297686; PubMed=12368813;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
  Read T.D., Eisen J.A., Seshadri R., Ward N., Meehe B., Clayton R.A.,
  Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
  DeBoy R.T., Dodson R.J., Durkin A.S., Hatt D.H., Kolony J.F.,
  Madupu R.T., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
  Vamathevan J., Weidman J., Impriali M., Lee K., Berry K., Lee C.,
  Mueller J., Khouri H., Gill J., Ureback T.R., McDonald L.A.,
  RA Feldblyum T.V., Smith H.O., Venter J.C., Nelson K.H., Fraser C.M.;
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
  Shewanella oneidensis.";
RL Nat. Biotechnol. 20:1118-1123(2002).
DR EMBL; AE015532; AAN53941.1; -.
DR TIGR; S00865; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 502 AA; 52441 MW; D08CA23D6C46B82D CRC64;

Query Match
  Best Local Similarity 15.7%; Score 122; DB 16; Length 502;
  Matches 41; Conservative 22; Mismatches 57; Indels 32; Gaps 5;

QY 29 GGNHNG-----GNN-----SSGPDSTLSIYOGSANA---ALALOS 61
  DB 231 GDNHTGFVVALAGSENDISMEQGSNNTAYLSMTTDDDTVDTTODGDSVTGSLIADI 290
QY 62 DAKSETTITGSGYGADVGAGADNSTIELTONGFRNNATIDOWNAKNYDQVTRVVT 121
  DB 291 QGDNDITTKOKDSDGAEFGWGDSDVDLKKRGDANRATFGATGTDNDPULSKGDN 350
QY 122 EAAHANQTRASDSSVWVRQVGFN---NATAN 149
  DB 351 ELV-APATGEDNSIEISQEGDANFAVATGN 381

RESULT 9

```

```

OBEIH3
ID OBEIH3 PRELIMINARY; PRT; 139 AA.
AC OBEIH3;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Minor curliin subunit CagB, putative.
GN S00866.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.
OX NCBI_TaxID=70863;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MR-1;
RX MEDLINE=22297686; PubMed=12368813;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
RA Weyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
RA Deboy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Vamathevan J., Weidman J., Umayam L.A., White O., Wolf A.M.,
RA Mueller J., Khouri H., Gill J., Uterback T.R., McDonald L.A.,
RA Feldblyum T.V., Smith H.O., Venter J.C., Nealsen K.H., Fraser C.M.,
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis."
RL Nat. Biotechnol. 20:1118-1123 (2002).
DR EMBL; AE015532; AAN53942.1; -.
DR TIGR; S00866; -.
KW Complete proteome.
SQ SEQUENCE 139 AA; 14811 MW; 41BC1CFA76957920 CRC64;

Query Match 14.1%; Score 109; DB 16; Length 139;
Best Local Similarity 30.1%; Pred. No. 0.28;
Matches 34; Conservative 19; Mismatches 46; Indels 14; Gaps 3;

QY 39 GSPDSTLSIYQGSANALALQSDARKSETTITQSGYNGADVGQADNSTIELTONGFR 98
DB 41 SGRNDLIDLVQOQTGQIVFGSGSDNS-AVYTOAGNDNISLVITQIGNNEVQLQVGAQ 99
QY 99 NNATIDQNAKNYQOLVTRVTHEMAHANOQASDSSVWVRQVGFNNATNOY 151
DB 100 NKASITQIGNDNLVOL-----NQIGS-GNFSIQIADGAISITQY 139

RESULT 10
ID Q7UWZ8 PRELIMINARY; PRT; 7716 AA.
AC Q7UWZ8;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN RB1661.
OS Rhodospirillum rubrum.
OC Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;
OC Planctomycetaceae; Pirellula.
OX NCBI_TaxID=117;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1;
RX MEDLINE=22735913; PubMed=12835416;
RA Gloeckner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,
RA Ludwig W., Gade D., Beck A., Borzym K., Heilmann K., Rabus R.,
RA Schleener H., Amann R., Reinhardt R.;
RT "Complete genome sequence of the marine planctomycete Pirellula sp.
RT strain 1."
RL Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303 (2003).
DR EMBL; BX294135; CAD72214.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 7716 AA; 797868 MW; D391A25BD96405CO CRC64;

Query Match 13.4%; Score 104; DB 16; Length 7716;

```

```

Best Local Similarity 29.1%; Pred. No. 81;
Matches 39; Conservative 18; Mismatches 57; Indels 20; Gaps 6;

QY 28 GGGNHN--GGGNS-----GGDSTLSIYQGSANALALQSDARKSETTITQSGY 75
DB 4048 GDNHGLTGGVAGSQEPTRSSSVGEGDTLSV--SVGGLLANDIEMDGDSTITTEIN- 4103
QY 76 GNGADVGQG---ADNSTIELTONGFRNNATIDQNAKNYQOLVTRVTHEMAHANOQASD 132
DB 4104 GNGSDVGNNTITLPSGATITQLEADGSPSYDPTIYQNLNGEATETFTYVDGNGNTDT 4163
QY 133 SSYMWKQVGRGNNNA 146
DB 4164 TSVALTITIG-GNDA 4176

RESULT 11
ID Q89D03 PRELIMINARY; PRT; 362 AA.
AC Q89D03;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical exported glutamine-rich protein.
GN BL17642.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RX MEDLINE=22484998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA Sasamoto S., Watanabe A., Ideasa K., Iriyuchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimpo S., Tsunoka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110."
RL DNA Res. 9:189-197 (2002).
DR EMBL; AP005962; BAC52907.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 362 AA; 39058 MW; D447E0D4E8433EC CRC64;

Query Match 13.3%; Score 103; DB 16; Length 362;
Best Local Similarity 31.6%; Pred. No. 2.6;
Matches 50; Conservative 14; Mismatches 80; Indels 14; Gaps 6;

QY 2 KLLKVAFAFAIVSGSALAGVVPQWGGGNGGNSGPDST---LSIYQGSANAL 57
DB 3 KRLFLATTAFAVAINTSALAQSSP---STNSNPSTTQROPDSTITSSSTPSSAQTNP 59
QY 58 ALQSDARKSETTITQSGYNGADVGQADNSTIELTONGFRNNATIDQNAKNYQDQ 113
DB 60 STNSAQOTSPSSSTQSAAGQTNSGTGT-NTTQAPTSNNSTNQOTSPSQNTAPSDQ 117
QY 114 LVTRVTHEMAH-ANQASDSSVWVRQVGFNNATNOY 150
DB 118 TQTNPTNRAQSANPPASGASQAQSPFGNSNTVTAQD 155

RESULT 12
ID Q19414 PRELIMINARY; PRT; 409 AA.
AC Q19414;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE F13B9.4 protein.
GN F13B9.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.

```

OX NCBI_TaxID=6239;
RN (1)
RP SEQUENCE FROM N.A.
RA McMurray A.A.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
RN (2)
RP SEQUENCE FROM N.A.
RX MEDLINE=9069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C. elegans: A platform for
investigating biology."
RL Science 282:2012-2016(1998).
DR EMBL; Z69383; CA93412.1; -.
DR FIR; T20847; T20847.
DR WormRep; F13E9.4; CE05606.
DR InterPro; IPR003677; Onchocerca_Ag.
DR Pfam; PF02520; DUF148; 1.
SQ SEQUENCE 409 AA; 43231 MW; B07DF0E4175C5739 CRC64;

Query Match 13.0%; Score 101; DB 5; Length 409;
Best Local Similarity 23.3%; Pred. No. 4.3;
Matches 37; Conservative 26; Mismatches 66; Indels 30; Gaps 5;

QY 15 SGGALGVVPQWGGGNNHNGGSGPDSLTIYQYGSANALALQSDARKSETTTTGG 74
DB 134 SGGAGSGSNNSFGGQGYGQNGFGQSGFGSGGSGNSLSANSNGNNNG--SSGG 191
QY 75 YGN--GADVGGADNSTIE--LTONGFRNN-----ATTIDWNAK 109
DB 192 YQNGRHRGQGGGSHSSSVSNNGYSSNGYSGNNNGPTSPFLANNSSAAQDYNNIV 251
QY 110 NYDQLVTRVTHEMAHANOASDSSVVRQVFGNNATA 148
DB 252 NNKSLTTNQINEQ---ASNWASANSVQAYIYETVRSR 287

RESULT 13
ID Q7X238 PRELIMINARY; PRT; 151 AA.
AC Q7X238;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Nucleation component of curlin monomers.
GN CSGB.
OS Enterobacter sakazakii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Enterobacter.
OX NCBI_TaxID=28141;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=FG39;
RA Zogaj X., Bokranz W., Nimtz M., Romling U.;
RT "Production of Cellulose and Curli Fimbriae by Members of the Family
Enterobacteriaceae Isolated from the Human Gastrointestinal Tract."
RL Infect. Immun. 72:4151-4158(2003).
RL EMBL; AJ515702; CAD56677.1; -.
SQ SEQUENCE 151 AA; 15985 MW; F08B2BD2A7882B7 CRC64;

Query Match 12.9%; Score 100; DB 2; Length 151;
Best Local Similarity 31.8%; Pred. No. 1.6;
Matches 35; Conservative 12; Mismatches 45; Indels 18; Gaps 4;

QY 10 AATVSGSALAGVVPQWGGGNNHNGGSGPDSLTIYQYGSANALALQSDARKSETT 69
DB 58 AATRGSGSLVSVSQ-----DAGNRAVD-----QSGTYNTAMIDQS--GNNDAG 103
QY 70 ITQSGYGAGADVQGGADNSTIELTONGFRNNATIDWNAKNDQLVTRVV 119
DB 104 ITQDYGNSAKTIQKSGNRRANITQYGTQTAIVVQ---KQSGMIRVI 149

RESULT 14

Q8EV84
ID Q8EV84 PRELIMINARY; PRT; 362 AA.
AC Q8EV84;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE P35 lipoprotein homolog.
GN MYP6840.
OS Mycoplasma penetrans.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=28227;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=HF-2;
RX MEDLINE=22354719; PubMed=1246555;
RA Sasaki Y., Ishikawa J., Yamashita A., Oshima K., Kenri T., Furuya K.,
Yoshino C., Horino A., Shiba T., Sasaki T., Hattori M.;
RT "The complete genomic sequence of Mycoplasma penetrans, an
intracellular bacterial pathogen in humans."
RL Nucleic Acids Res. 30:5293-5300(2002).
RL EMBL; AP004172; BAC44476.1; -.
DR InterPro; IPR000437; Prok Lipoprot. S.
DR ProSITE; PS00013; PROKAR Lipoprotein; 1.
KW lipoprotein; Complete proteome.
SQ SEQUENCE 362 AA; 38547 MW; 8DA27F70D19D354F CRC64;

Query Match 12.9%; Score 100; DB 16; Length 362;
Best Local Similarity 23.9%; Pred. No. 4.5;
Matches 42; Conservative 33; Mismatches 63; Indels 38; Gaps 8;

QY 1 MRLKVAFAIIVSGS--ALAGVP-----QWGGGNNHNGGSGSG----PDS 43
DB 1 MKIKKIKLKALALGAFGIVATVPYIVSSCSTSDNNNGNNNNNGQDGGCQOQTEI 60
QY 44 TLSIYQYGSANALALQSDARKSETTTTGG-----YGNADVQGGADNSTIEL---- 92
DB 61 TPTIKKSVLSGALSKTYIPANKSTSDLIADIKANPTNFDGEALKDILKATVSVNG 120
QY 93 -TONGFRNNATIDWNAKNDQLVTRVTHEMAHANOASDSSV--MVRQVFGNN 145
DB 121 FIESFTKGD-TYETWSAKYDKKGT-----YAGNSKQIDISINDLETQLGDSNN 169

RESULT 15
ID Q89J13 PRELIMINARY; PRT; 171 AA.
AC Q89J13;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE CsgA protein.
GN CSGA OR BL5300.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiales; Bradyrhizobium.
OX NCBI_TaxID=375;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RX MEDLINE=22484998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA Sasamoto S., Watanabe A., Idegawa K., Iriyuchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimpo S., Teurloka H., Wada T., Yamada M.,
RA Tabara S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
Bradyrhizobium japonicum USDA110."
RL DNA Res. 9:189-197(2002).
RL EMBL; AP005954; BAC50565.1; -.
KW Complete proteome.
SQ SEQUENCE 171 AA; 17448 MW; 995DB08C01498381 CRC64;

Query Match 12.8%; Score 99.5; DB 16; Length 171;
Best Local Similarity 25.8%; Pred. No. 2;

Matches 40; Conservative 21; Mismatches 63; Indels 31; Gaps 4;

QY 1 MKLLKVAFAIYVSGSALAGVPPQWGGGNNHGGNSGPDSTLSTIYQYGSANALALQ 60
 Db 40 MKRLFPASVAVIALSSAAQA-----NTSTVQVGLVNGSSVTQ 78
 QY 61 SPARKSETTITQSGYNGADVGQAD---NSTIELTONGFRNNATIDQWNAKNYDQVLT 116
 Db 79 NGILTNDSSSTTQIGLNGASTWQGTSSPSLNNVSTVNAQGVONSALTGTGVAFGNNGSAIT 138
 QY 117 RYVTHMAHANQTAQDSSVMVRQVFG-NNATANQ 150
 Db 139 QNSFGPPALQNNNSA-----VGQLSRGINTSVSQ 168

Search completed: March 11, 2004, 18:40:45
 Job time : 32.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:13:53 ; Search time 45.9 Seconds
(without alignment)
929.514 Million cell updates/sec

Title: US-09-543-407-26

Perfect score: 782
Sequence: 1 MKLKVAAFAALVSGSALA.....DSSVMVROVFGNNATANQY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*

1: geneseqp19808:*

2: geneseqp19908:*

3: geneseqp20008:*

4: geneseqp20018:*

5: geneseqp20028:*

6: geneseqp20038:*

7: geneseqp20038:*

8: geneseqp20048:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	782	100.0	151	3	AAB36353 Aab36353 Agfa::PT3
2	709	90.7	151	3	AAB36350 Aab36350 Agfa::PT3
3	692	88.5	151	2	AAB74625 Aab74625 Agfa sequ
4	682	88.5	151	3	AAB36341 Aab36341 Salmonell
5	687	87.9	151	2	AAB23570 Aab23570 Salmonell
6	675	86.3	151	3	AAB36349 Aab36349 Agfa::PT3
7	617	78.9	151	3	AAB36354 Aab36354 Agfa::PT3
8	614	78.5	151	3	AAB36346 Aab36346 Agfa::PT3
9	612	78.3	151	3	AAB36347 Aab36347 Agfa::PT3
10	609	77.9	151	3	AAB36352 Aab36352 Agfa::PT3
11	601	76.9	151	3	AAB36351 Aab36351 Agfa::PT3
12	600	76.7	151	3	AAB36355 Aab36355 Agfa::PT3
13	577	73.8	151	3	AAB36348 Aab36348 Agfa::PT3
14	523	66.9	151	3	AAB36343 Aab36343 Escherich
15	518	66.2	151	7	ABR82651 ABR82651 E. coli C
16	507	64.8	120	2	AAB2761 Aab2761 Agfa sequ
17	507	64.8	120	2	AAB23559 Aab23559 Salmonell
18	445	56.9	142	2	AAB32664 Aab32664 Fibronect
19	373	47.7	122	2	AAB2663 Aab2663 FMB curli
20	188	24.0	45	3	AAB36316 Aab36316 Salmonell
21	132	16.9	22	3	AAB36318 Aab36318 Salmonell
22	123	15.7	23	3	AAB36331 Aab36331 Salmonell
23	123	15.7	23	3	AAB36336 Aab36336 Salmonell
24	123	15.7	23	3	AAB36338 Aab36338 Salmonell
25	113	14.5	24	7	ABR82644 ABR82644 E. coli c

26	112	14.3	151	3	AAB36344 Aab36344 Escherich
27	111	14.2	22	3	AAB36322 Aab36322 Salmonell
28	111	14.2	22	3	AAB36327 Aab36327 Salmonell
29	111	14.2	22	3	AAB36337 Aab36337 Salmonell
30	107	13.7	151	3	AAB36342 Aab36342 Salmonell
31	105	13.4	597	4	AAB08231 Aab08231 Polypepti
32	102	13.0	26	7	ABR82649 ABR82649 E. coli V
33	100	12.8	943	2	AAB64378 Aab64378 Mycobacte
34	100	12.8	943	2	AAB81745 Aab81745 M. tuberc
35	100	12.8	943	2	AAB39032 Aab39032 M. tuberc
36	100	12.8	943	2	AAB39175 Aab39175 M. tuberc
37	100	12.8	3300	6	ABU36445 Aab36445 Protein e
38	98.5	12.6	738	2	AAB56163 Aab56163 New DNA 8
39	97.5	12.5	850	4	ABB65764 Aab65764 Drosophila
40	97.5	12.5	1028	4	ABD62708 Aab62708 Arabidops
41	96.5	12.3	447	3	AAG29728 Aab29728 Arabidops
42	96.5	12.3	468	3	AAG29727 Aab29727 Arabidops
43	96	12.3	19	3	AAB36323 Aab36323 Salmonell
44	96	12.3	19	3	AAB36336 Aab36336 Salmonell
45	96	12.3	19	3	AAB36328 Aab36328 Salmonell

ALIGNMENTS

RESULT 1
AAB36353
ID AAB36353 standard; protein; 151 AA.
XX
AC AAB36353;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa::PT3#8 amino acid sequence SEQ ID NO:26.
XX
KW Salmonella; agfa, chromosomal gene replacement; fimbria; epitope;
KW vaccine; immune response; immunogen.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
XX 12-OCT-2000.
XX PD
XX 05-APR-2000; 2000WO-CA000356.
XX PR
XX 05-APR-1999; 99US-0127888P.
XX PA (UYVI-) UNIV VICTORIA.
XX PI White AP, Doran JL, Collison SK, Kay W;
XX WPI; 2000-672631/65.
XX DR N-PSDB; AAC64629.
XX PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
XX which encodes foreign epitope or antigen, expresses recombinant Agfa
XX protein useful for eliciting immune response in animal.
XX PS Disclosure; Page 138; 139pp; English.
XX
XX The present invention describes a recombinant agfa gene (1) where a
XX segment of the gene has been replaced by a segment of a foreign DNA
XX sequence which encodes a foreign epitope or antigen. Also described are:
XX (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
XX assembly system of strains of Salmonella, Escherichia coli and
XX Enterobacteriaceae for the production of fimbriae comprising recombinant
XX Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
XX directing recombination of a recombinant gene into the chromosome of the
XX homologous species; (3) directing recombination of a recombinant gene
XX back into the chromosome of the homologous species, replacing the native

CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fibrillar presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fibrin protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fibrillar subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fibrillae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

XX Sequence 151 AA;

Query Match 100.0%; Score 782; DB 3; Length 151;
 Best Local Similarity 100.0%; Pred. No. 2.3e-67;
 Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNHNGGNSGPDSTLSIYQGSANALAYDQ 60
 DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNHNGGNSGPDSTLSIYQGSANALAYDQ 60
 QY 61 LVTRVVTHEMAHAGYNGADVGQGDNSTIELTNGFRNNATTIOMNAKNSDITVQYGG 120
 DB 61 LVTRVVTHEMAHAGYNGADVGQGDNSTIELTNGFRNNATTIOMNAKNSDITVQYGG 120
 QY 121 NNAAALVNOQTASDSVWVRQVGFNNATANQY 151
 DB 121 NNAAALVNOQTASDSVWVRQVGFNNATANQY 151

RESULT 2
 AAB36350
 ID AAB36350 standard; protein; 151 AA.

XX AAB36350;
 AC
 XX 26-FEB-2001 (first entry)
 DT
 XX
 DE Agfa::PT3#5 amino acid sequence SEQ ID NO:20.
 XX
 KW *Salmonella*; agfa; chromosomal gene replacement; fibrin; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS *Salmonella enteritidis*.
 OS *Escherichia coli*.
 OS Synthetic.

XX
 PN WO200060102-A2.
 PD 12-OCT-2000.
 PD 05-APR-2000; 2000WO-CA000356.
 PF 05-APR-1999; 99US-0127888P.
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 PI White AP, Doran JL, Collison SK, Kay WW,
 DR WPI: 2000-672631/65.
 DR N-PSDB; AAC64626.

XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 137; 139pp; English.

XX The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fibrillae (SEF17/RAF) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC Enterobacteriaceae for the production of fibrillae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fibrin subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fibrillar presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fibrin protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fibrillar subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fibrillae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

XX Sequence 151 AA;

Query Match 90.7%; Score 709; DB 3; Length 151;
 Best Local Similarity 89.9%; Pred. No. 2.5e-60;
 Matches 143; Conservative 0; Mismatches 0; Indels 16; Gaps 2;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNHNGGNSGPDSTLSIYQGSANALAYDQ 57
 DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNHNGGNSGPDSTLSIYQGSANALAYDQ 60
 QY 58 -----YDQLVTRVVTHEMAHAGYNGADVGQGDNSTIELTNGFRNNATTIOMNAKNSD 112
 DB 61 SDARKYDQLVTRVVTHEMAHAGYNGADVGQGDNSTIELTNGFRNNATTIOMNAKNSD 112
 QY 113 ITVQYGGNNALVNOQTASDSVWVRQVGFNNATANQY 151
 DB 113 ITVQYGGNNALVNOQTASDSVWVRQVGFNNATANQY 151

RESULT 3
 AAR74625
 ID AAR74625 standard; protein; 151 AA.

XX AAR74625;
 AC
 XX 25-MAR-2003 (revised)
 DT 26-JUN-1995 (first entry)
 XX
 DE Agfa sequence.

XX
 KW *Salmonella*; Agfa; vaccine.
 KW *Salmonella*.
 OS *Salmonella*.
 OS WO9425598-A2.
 PN 10-NOV-1994.
 PD 26-APR-1994; 94WO-IB000207.
 PF 26-APR-1994; 94WO-IB000207.
 PR 26-APR-1993; 93US-00054452.
 XX
 PA (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
 XX (KING/) KING J.

PI Kay WM, Collinson SK, Clouthier SC, Doran JL;
 DR MPI; 1994-358275/44.
 DR N-PSDB; AAO87467.
 XX
 PT Eliciting an immune response to Salmonella - using attenuated Salmonella
 PT strains, vector constructs, or compans. contg. fimbrial type proteins.
 XX
 PS Disclosure; Fig 7B; 95pp; English.
 CC The Salmonella Agfa protein and DNA are used in vaccine and genetic
 CC immunization compositions, respectively, to elicit an immune response to
 CC Salmonella in animals (e.g. food producing animals) and humans. (Updated
 CC on 25-MAR-2003 to correct PW field.)
 CC
 SQ Sequence 151 AA;
 Query Match 88.5%; Score 692; DB 2; Length 151;
 Best Local Similarity 90.7%; Pred. No. 1.1e-58;
 Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;
 QY 1 MLKLKVAAPFAIIVSGSALAGVVPQWGGGNNHGGSSGPDSTLSIYQGSANALAYDQ 60
 DB 1 MLKLKVAAPFAIIVSGSALAGVVPQWGGGNNHGGSSGPDSTLSIYQGSANALAYDQ 60
 QY 61 LVTRVVTHEMAHAGVNGADVGQAGDNSTIELTONGFRNNATIDQWNAKNSDITVGOYGS 120
 DB 61 SDARSETTITQSGYNGADVGQAGDNSTIELTONGFRNNATIDQWNAKNSDITVGOYGS 120
 QY 121 NNAALVNOQTASDSSVMVROVGFGNNTANQY 151
 DB 121 NNAALVNOQTASDSSVMVROVGFGNNTANQY 151
 RESULT 4
 AAB36341
 ID AAB36341 standard; protein; 151 AA.
 AC AAB36341;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Salmonella enteritidis Agfa amino acid sequence SEQ ID NO:5.
 XX
 KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS
 PN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UUYI-) UNIV VICTORIA.
 PI White AP, Doran JL, Collinson SK, Kay WM;
 DR MPI; 2000-672631/65.
 DR N-PSDB; AAC64617.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 135; 139pp; English.
 CC The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:

CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, Caga and Agfa-homologue fimbria subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, B. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 CC
 SQ Sequence 151 AA;
 Query Match 88.5%; Score 692; DB 3; Length 151;
 Best Local Similarity 90.7%; Pred. No. 1.1e-58;
 Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;
 QY 1 MLKLKVAAPFAIIVSGSALAGVVPQWGGGNNHGGSSGPDSTLSIYQGSANALAYDQ 60
 DB 1 MLKLKVAAPFAIIVSGSALAGVVPQWGGGNNHGGSSGPDSTLSIYQGSANALAYDQ 60
 QY 61 LVTRVVTHEMAHAGVNGADVGQAGDNSTIELTONGFRNNATIDQWNAKNSDITVGOYGS 120
 DB 61 SDARSETTITQSGYNGADVGQAGDNSTIELTONGFRNNATIDQWNAKNSDITVGOYGS 120
 QY 121 NNAALVNOQTASDSSVMVROVGFGNNTANQY 151
 DB 121 NNAALVNOQTASDSSVMVROVGFGNNTANQY 151
 RESULT 5
 AAM23570
 ID AAM23570 standard; protein; 151 AA.
 AC AAM23570;
 XX
 DT 25-MAR-2003 (revised)
 DT 29-SEP-1997 (first entry)
 XX
 DE Salmonella enteritidis 27655-3b agfa.
 XX
 KW Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody.
 XX
 OS Salmonella enteritidis.
 OS
 PN US5635617-A.
 XX
 PD 03-JUN-1997.
 XX
 PF 26-APR-1994; 94US-00233788.
 XX
 PR 26-APR-1993; 93US-00054452.
 XX
 PA (UUYI-) UNIV VICTORIA INNOVATION & DEV CORP.
 PI Collinson SK, Kay WM, Doran JL;

XX WPI, 1997-309886/28.
DR N-PSDB; AAT74142.
XX
PT Isolated *Salmonella* gene agfa - used for diagnosis of *Salmonella* or
XX enteropathogenic bacteria of the Enterobacteria family.
PS
XX Example 2; Fig 7; 85pp; English.
CC The present sequence represents agfa encoded by the full agfa gene
CC derived from *Salmonella enteritidis* 27655-3b. The nucleic acid can be
CC used to provide diagnostic assays for *Salmonella* and/or enteropathogenic
CC bacteria of the family Enterobacteria. It can also be used to provide
CC proteins and antibodies which can be used for assays. The nucleic acid
CC sequence can be used to provide probes or primers which can specifically
CC hybridise to nucleic acid molecules from greater than 99% of *Salmonella*
CC strains that are pathogenic to warm-blooded animals relative to nucleic
CC acid molecules from virtually all other microbial organisms. (Updated on
CC 25-MAR-2003 to correct PF field.)
SQ
XX
XX Sequence 151 AA;
Query Match 87.9%; Score 687; DB 2; Length 151;
Best Local Similarity 90.1%; Pred. No. 3.3e-58;
Matches 136; Conservative 2; Mismatches 13; Indels 0; Gaps 0;
QY 1 MKLTKVAFPAIIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLTIYQGSANALAYDQ 60
DB 1 MKLTKVAFPAIIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLTIYQGSANALALQ 60
QY 61 LVTRVVTHEMAHAGYGADVGAGADNSTIELTQNGFRNNATTIDQWNAKNSDITVGQYGG 120
DB 61 SDAKSETTITQSGYGAGADVGAGADNSTIELTQNGFRNNATTIDQWNAKNSDITVGQYGG 120
QY 121 NNAALVNQTASDSSVMVRQVGFNNATTANQY 151
DB 121 NNPAALVNQTASDSSVMVRQVGFNNATTANQY 151
RESULT 6
AAB36349
ID AAB36349 standard; protein; 151 AA.
AC AAB36349;
XX
XX 26-FEB-2001 (first entry)
DT
XX
XX Agfa::PT3#4 amino acid sequence SEQ ID NO:18.
DE
XX
XX *Salmonella*; agfa; chromosomal gene replacement; fimbria; epitope;
KW vaccine; immune response; immunogen.
XX
XX *Salmonella enteritidis*.
OS
XX *Escherichia coli*.
OS
XX Synthetic.
OS
XX WO200060102-A2.
PN
XX
XX 12-OCT-2000.
PD
XX
XX 05-APR-2000; 2000WO-CA000356.
PF
XX
XX 05-APR-1999; 99US-0127888P.
PR
XX
XX (UYVI-) UNIV VICTORIA.
PA
XX
XX White AP, Doran JL, Collison SK, Kay WW,
PI
XX
XX WPI; 2000-672631/65.
DR
XX
XX N-PSDB; AAC64625.
DR
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa

PT protein useful for eliciting immune response in animal.
XX
XX
PS Disclosure; Page 136; 139pp; English.
XX
XX The present invention describes a recombinant agfa gene (1) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF7/7AR) nucleation depended
CC assembly system or strains of *Salmonella*, *Escherichia coli* and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbria presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbria subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 151 AA;
Query Match 86.3%; Score 675; DB 3; Length 151;
Best Local Similarity 81.9%; Pred. No. 4.7e-57;
Matches 136; Conservative 0; Mismatches 0; Indels 30; Gaps 2;
QY 1 MKLTKVAFPAIIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLTIYQGSANALAYDQ 60
DB 1 MKLTKVAFPAIIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLTIYQGSANALALQ 45
QY 61 LVTRVVTHEMAHA-----GYGAGADVGAGADNSTIELTQNGFRNNATTIDQ 105
DB 46 LVTRVVTHEMAHALSDAKSETTITQSGYGAGADVGAGADNSTIELTQNGFRNNATTIDQ 105
QY 106 WNAKNSDITVGQYGGNNALVNQTASDSSVMVRQVGFNNATTANQY 151
DB 106 WNAKNSDITVGQYGGNNALVNQTASDSSVMVRQVGFNNATTANQY 151
RESULT 7
AAB36354
ID AAB36354 standard; protein; 151 AA.
AC AAB36354;
XX
XX 26-FEB-2001 (first entry)
DT
XX
XX Agfa::PT3#9 amino acid sequence SEQ ID NO:28.
DE
XX
XX *Salmonella*; agfa; chromosomal gene replacement; fimbria; epitope;
KW vaccine; immune response; immunogen.
XX
XX *Salmonella enteritidis*.
OS
XX *Escherichia coli*.
OS
XX Synthetic.
OS
XX WO200060102-A2.
PN
XX
XX 12-OCT-2000.
PD
XX
XX 05-APR-2000; 2000WO-CA000356.
PF
XX
XX

PR 05-APR-1999; 99US-0127888P.
 XX (UYVI-) UNIV VICTORIA.
 PA White AP, Doran JL, Collison SK, Kay WM;
 PI WPI, 2000-672631/65.
 XX N-PSDB; AAC64630.
 DR
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 138; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/7AF) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;
 XX
 Query Match 78.9%; Score 617; DB 3; Length 151;
 Best Local Similarity 73.6%; Pred. No. 1.8e-51;
 Matches 18; Conservative 0; Mismatches 0; Indels 46; Gaps 2;
 QY 1 MLLKVAAPAAIVSGSALAGVVPWGGGNNHNGGNSGPDSTLSIYOGSANAAL 57
 DB 1 MLLKVAAPAAIVSGSALAGVVPWGGGNNHNGGNSGPDSTLSIYOGSANAAL 60
 QY 58 -----YDQLVTRVVTTHMAAGYNGADYGGADNSTLTQNGF 97
 DB 61 SPARKSETTITOSGYNGADYDQLVTRVVTTHMAA-----F 97
 QY 98 RNNATIDOMNAKNSDITVGOYGNNALVNOTASDSSVWVROVFGNNAATANOY 151
 DB 98 RNNATIDOMNAKNSDITVGOYGNNALVNOTASDSSVWVROVFGNNAATANOY 151
 RESULT 8
 ID AAB36346 standard; protein; 151 AA.
 XX AAB36346;
 XX 26-FEB-2001 (first entry)
 DT
 XX Agfa::PT3#1 amino acid sequence SEQ ID NO:12.
 DE
 XX *Salmonella*; agfa; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.

XX *Salmonella enteritidis*.
 OS *Escherichia coli*.
 OS Synthetic.
 XX WO200060102-A2.
 PN 12-OCT-2000.
 PD
 XX 05-APR-2000; 2000WO-CA000356.
 PF
 XX 05-APR-1999; 99US-0127888P.
 PR (UYVI-) UNIV VICTORIA.
 XX
 PA White AP, Doran JL, Collison SK, Kay WM;
 PI WPI, 2000-672631/65.
 XX N-PSDB; AAC64622.
 DR
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 135; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/7AF) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;
 XX
 Query Match 78.5%; Score 614; DB 3; Length 151;
 Best Local Similarity 80.8%; Pred. No. 3.5e-51;
 Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;
 QY 1 MLLKVAAPAAIVSGSALAGVVPWGGGNNHNGGNSGPDSTLSIYOGSANAAL 60
 DB 1 MLLKVAAPAAIVSGSALAGVVPWGGGNNHNGGNSGPDSTLSIYOGSANAAL 60
 QY 61 LVTRVVTTHMAAGYNGADYGGADNSTLTQNGFRNATIDOMNAKNSDITVGOYGG 120
 DB 61 SPARKSETTITOSGYNGADYGGADNSTLTQNGFRNATIDOMNAKNSDITVGOYGG 120
 QY 121 NNAALVNOTASDSSVWVROVFGNNAATANOY 151
 DB 121 NNAALVNOTASDSSVWVROVFGNNAATANOY 151
 RESULT 9

AAB6347
 ID AAB6347 standard; protein; 151 AA.
 AC AAB6347;
 XX
 DT 26-FEB-2001 (first entry)
 DE Agfa::PT3#2 amino acid sequence SEQ ID NO:14.
 XX
 KW Salmoneila: agfa; chromosomal gene replacement; fimbriin; epitope;
 KM vaccine; immune response; immunogen.
 OS Salmoneila enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 PN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay WM;
 XX
 DR WPI: 2000-672631/65.
 DR N-PSDB; AAC64623.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 PS
 PS Disclosure; Page 136; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SFP17/TAF) nucleation depended
 CC assembly system of strains of Salmoneila, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbriin subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmoneila, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbriin protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;
 Query Match 78.3%; Score 612; DB 3; Length 151;
 Best Local Similarity 81.5%; Pred. No. 5.5e-51;
 Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPOMGGGNNHGGSSGPDSTLTSTYOGSANAALYDQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPOMGGGNNHGGSSGPDSTLTSTYOGSANAALALQ 60

QY 61 LVTRVVTHEMAHAGYGNAGADYGGADNSTLTETLONGFNNATTDOWNAKSDITGOYGS 120
 DB 61 SDARKSETTTTGGYGADYGGADNSTLTETLONGFNNATTDOWNAKSDITGOYDQ 120
 QY 121 NNAALVNOTASDSSVWVRQYGFNNATANQY 151
 DB 121 LVTRVVTHEMAHASVWVRQYGFNNATANQY 151
 RESULT 10
 AAB6352
 ID AAB6352 standard; protein; 151 AA.
 AC AAB6352;
 XX
 DT 26-FEB-2001 (first entry)
 DE Agfa::PT3#7 amino acid sequence SEQ ID NO:24.
 XX
 KW Salmoneila: agfa; chromosomal gene replacement; fimbriin; epitope;
 KM vaccine; immune response; immunogen.
 OS Salmoneila enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 PN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay WM;
 XX
 DR WPI: 2000-672631/65.
 DR N-PSDB; AAC64628.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 PS
 PS Disclosure; Page 138; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SFP17/TAF) nucleation depended
 CC assembly system of strains of Salmoneila, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbriin subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmoneila, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbriin protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in

CC the exemplification of the present invention
XX Sequence 151 AA;
SQ

Query Match 77.9%; Score 609; DB 3; Length 151;
Best Local Similarity 82.1%; Pred. No. 1.1e-50;
Matches 124; Conservative 4; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAAPFAIVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYOGSANAALYDQ 60
DB 1 MKLLKVAAPFAIVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYOGSANAALYDQ 60
QY 61 LVTRVVTHEMAHAGYNGADVGQADNSTIELTONGFRNNATTIDQNAKNSDITVGQYGG 120
DB 61 SPARKSETTITQSGYNGADVGQADNSTIELTONGFRNNATTIDQNAKNSDITVGQYGG 120
QY 121 NNAALVNOTASDSSVWVROVGFNNATANOY 151
DB 121 NNAALVNOTASDSSVWVROVGFNNATANOY 151

RESULT 11
AAB36351
ID AAB36351 standard; protein; 151 AA.
XX
AC AAB36351;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa: PT3#6 amino acid sequence SEQ ID NO:22.
XX
KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
XX vaccine; immune response; immunogen.
XX
OS Salmonella enteritidis.
XX
OS Escherichia coli.
XX
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000MO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UUYI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WM;
XX
DR MPI; 2000-672631/65.
XX
DR N-PSDB; AAC64627.
XX
PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 137; 139pp; English.
XX
CC The present invention describes a recombinant agfa gene (1) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/7AP) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, Caga and Agfa-homologue fimbria subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or

CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbria subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention

SQ Sequence 151 AA;
Query Match 76.9%; Score 601; DB 3; Length 151;
Best Local Similarity 81.5%; Pred. No. 6.3e-50;
Matches 123; Conservative 4; Mismatches 24; Indels 0; Gaps 0;

QY 1 MKLLKVAAPFAIVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYOGSANAALYDQ 60
DB 1 MKLLKVAAPFAIVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYOGSANAALYDQ 60
QY 61 LVTRVVTHEMAHAGYNGADVGQADNSTIELTONGFRNNATTIDQNAKNSDITVGQYGG 120
DB 61 SPARKSETTITQSGYNGADVGQADNSTIELTONGFRNNATTIDQNAKNSDITVGQYGG 120
QY 121 NNAALVNOTASDSSVWVROVGFNNATANOY 151
DB 121 NNAALVNOTASDSSVWVROVGFNNATANOY 151

RESULT 12
AAB36355
ID AAB36355 standard; protein; 151 AA.
XX
AC AAB36355;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa: PT3#10 amino acid sequence SEQ ID NO:30.
XX
KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
XX vaccine; immune response; immunogen.
XX
OS Salmonella enteritidis.
XX
OS Escherichia coli.
XX
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000MO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UUYI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WM;
XX
DR MPI; 2000-672631/65.
XX
DR N-PSDB; AAC64631.
XX
PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 139; 139pp; English.
XX
CC The present invention describes a recombinant agfa gene (1) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:

CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant AgfA
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

CC Sequence 151 AA;

Query Match 76.7%; Score 600; DB 3; Length 151;
 Best Local Similarity 81.5%; Pred. No. 7.8e-50;
 Matches 123; Conservative 4; Mismatches 24; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNNHNGSGSPDSTLSIYQGSANAAALYDQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNNHNGSGSPDSTLSIYQGSANAAALYDQ 60
 QY 61 LVTRVVTHEMAHAGYGNAGADVCGADNSTITELTONGFRNNATITDWMKNSDITVGYGCG 120
 DB 61 SDARKSETTITQSGYGNGADVCGADNSTITELTONGFRNNATITDVLTRVVTHEMAHAG 120
 QY 121 NNAALVNOTASDSVYWRQVGFNNATANY 151
 DB 121 NNAALVNOTASDSVYWRQVGFNNATANY 151

RESULT 13
 AAB36348
 ID AAB36348 standard; protein; 151 AA.

AC AAB36348;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE AgfA::PT3#3 amino acid sequence SEQ ID NO:16.
 XX
 KW *Salmonella*; agfA; chromosomal gene replacement; fimbrial; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS *Salmonella enteritidis*.
 OS *Escherichia coli*.
 OS Synthetic.
 PN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay WW,
 XX WPI; 2000-672631/65.
 DR

DR N-PSDB; AAC64624.
 XX
 PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant AgfA
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 136; 139pp; English.

CC The present invention describes a recombinant agfA gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant AgfA
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

CC Sequence 151 AA;

Query Match 73.8%; Score 577; DB 3; Length 151;
 Best Local Similarity 80.8%; Pred. No. 1.3e-47;
 Matches 122; Conservative 5; Mismatches 24; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNNHNGSGSPDSTLSIYQGSANAAALYDQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNNHNGSGSPDSTLSIYQGSANAAALYDQ 60
 QY 61 LVTRVVTHEMAHAGYGNAGADVCGADNSTITELTONGFRNNATITDWMKNSDITVGYGCG 120
 DB 61 SDARKSETTITQSGYGNGADVCGADNSTITELTONGFRNNATITDWMKNSDITVGYGCG 120
 QY 121 NNAALVNOTASDSVYWRQVGFNNATANY 151
 DB 121 NNAALVNOTASDSVYWRQVGFNNATANY 151

RESULT 14
 AAB36343

ID AAB36343 standard; protein; 151 AA.

AC AAB36343;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE *Escherichia coli* CsgA amino acid sequence SEQ ID NO:7.
 XX
 KW *Salmonella*; agfA; chromosomal gene replacement; fimbrial; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS *Escherichia coli*.
 OS Synthetic.
 PN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX

PF 05-APR-2000; 2000MO-C4000356.
XX
XX 05-APR-1999; 99US-0127888P.
XX
XX (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WM;
XX
XX MPI; 2000-672631/65.
DR N-PSDB; AAC64619.
XX
PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 135; 139pp; English.
XX
CC The present invention describes a recombinant agfa gene (1) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of *Salmonella*, *Escherichia coli* and
CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal.
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
CC *Enterobacteriaceae* host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 151 AA;
XX
Query Match 66.9%; Score 523; DB 3; Length 151;
Best Local Similarity 68.9%; Pred. No. 2e-42;
Matches 104; Conservative 18; Mismatches 29; Indels 0; Gaps 0;
QY 1 MLLKLVAAFAAIVVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYQGSANALAYDQ 60
DB 1 MLLKLVAAFAAIVVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYQGSANALAYDQ 60
QY 61 LVTRVVTHEMAHAGYNGADVQGGADNSTIELTQNGFRNNATIDQNNAKNSDITVQYGG 120
DB 61 TDARNSDLITTHQGGNGADVQGGSDSSIDLTRGFGNSATIDQNNKNSDITVQYGG 120
QY 121 NNAALVNOTASDSSVWVROYGFGNNAATANOY 151
DB 121 GNGAAVDQTAHSNSVWVTVQVFGNNAATANOY 151
XX
RESULT 15
ID ABR82651
AC ABR82651 standard; proteoin; 151 AA.
XX
XX ABR82651;
DT 04-DEC-2003 (first entry)
XX
DE E. coli CsgA subunit 15 kDa proteoin.
XX

KW Plasma proteoin; immune response; antibacterial; vaccine; gene therapy.
XX
XX *Escherichia coli*.
OS
XX
PN MO2003064446-A2.
XX
XX 07-AUG-2003.
PD
XX
XX 30-JAN-2003; 2003WO-EP000943.
PF
XX
PR 31-JAN-2002; 2002GB-00002275.
XX
XX (HANS-) HANSA MEDICAL RES AB.
XX
PI Bjoerck L, Olsen A, Wikstroem M, Herwald H;
XX
XX MPI; 2003-646136/61.
DR N-PSDB; ACF36153.
XX
XX
XX New isolated peptide capable of binding a mammalian plasma proteoin,
PT useful in the manufacture of a medicament for the prevention and/or
PT treatment of a bacterial infection, such as *Escherichia coli*, *Salmonella*
PT or *Shigella* infections.
XX
PS Disclosure; Page 41-42; 42pp; English.
XX
XX
XX The invention relates to an isolated peptide capable of binding a
XX mammalian plasma proteoin or of generating an immune response in a mammal
XX selected from sequences shown in ABR82642, ABR82646-49. The peptide or
XX antibody is useful for treating a bacterial infection in a human or
XX animal or in the manufacture of a medicament for the prophylactic
XX treatment of a bacterial infection, such as *Escherichia coli*, *Salmonella*
XX or *Shigella* infection. The peptide that is immobilized on a solid support
XX is also useful as a reagent for determining the ability of a plasma
XX protein to bind to bacteria. The present sequence represents an *E. coli*
XX 15 kDa proteoin
XX
SQ Sequence 151 AA;
XX
Query Match 66.2%; Score 518; DB 7; Length 151;
Best Local Similarity 68.2%; Pred. No. 6.2e-42;
Matches 103; Conservative 18; Mismatches 30; Indels 0; Gaps 0;
QY 1 MLLKLVAAFAAIVVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYQGSANALAYDQ 60
DB 1 MLLKLVAAFAAIVVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYQGSANALAYDQ 60
QY 61 LVTRVVTHEMAHAGYNGADVQGGADNSTIELTQNGFRNNATIDQNNAKNSDITVQYGG 120
DB 61 TDARNSDLITTHQGGNGADVQGGSDSSIDLTRGFGNSATIDQNNKNSDITVQYGG 120
QY 121 NNAALVNOTASDSSVWVROYGFGNNAATANOY 151
DB 121 GNGAAVDQTAHSNSVWVTVQVFGNNAATANOY 151
XX
Search completed: March 11, 2004, 18:33:45
Job time : 46.9 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:25:54 ; Search time 12.4 Seconds
(without alignments)
628.671 Million cell updates/sec

Title: US-09-543-407-26

Perfect score: 782

Sequence: 1 MLKLKVAAPALIVSGSALA.....DSVMYQVFGNNATANYQ 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: /cgnt2_6/prodata/2/1aa/5A-COMB.pep:*
2: /cgnt2_6/prodata/2/1aa/5B-COMB.pep:*
3: /cgnt2_6/prodata/2/1aa/6A-COMB.pep:*
4: /cgnt2_6/prodata/2/1aa/6B-COMB.pep:*
5: /cgnt2_6/prodata/2/1aa/PCITUS-COMB.pep:*
6: /cgnt2_6/prodata/2/1aa/backfillset1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	687	87.9	151	1 US-08-233-788A-59	Sequence 59, Appl
2	507	64.8	120	1 US-08-233-788A-57	Sequence 57, Appl
3	100	12.8	943	4 US-09-056-556-204	Sequence 204, App
4	100	12.8	943	4 US-09-072-596-199	Sequence 199, App
5	100	12.8	943	4 US-09-477-135A-131	Sequence 131, App
6	100	12.8	943	4 US-09-072-967-204	Sequence 204, App
7	98.5	12.6	728	3 US-08-864-038A-3	Sequence 3, Appl
8	93	11.9	82	4 US-09-336-447A-5	Sequence 5, Appl
9	91.5	11.7	975	4 US-09-328-352-4764	Sequence 4764, Ap
10	88	11.3	273	4 US-09-328-352-6167	Sequence 6167, Ap
11	86	11.0	1415	4 US-09-252-991A-26438	Sequence 26438, A
12	82	10.5	673	3 US-09-196-387-8	Sequence 8, Appl
13	82	10.5	673	4 US-09-841-835-8	Sequence 8, Appl
14	82	10.5	949	3 US-09-196-387-10	Sequence 10, Appl
15	82	10.5	949	4 US-09-841-835-10	Sequence 10, Appl
16	82	10.5	1337	3 US-09-196-387-2	Sequence 2, Appl
17	82	10.5	1337	4 US-09-841-835-2	Sequence 2, Appl
18	82	10.5	1337	4 US-09-972-115A-8	Sequence 8, Appl
19	81.5	10.4	873	4 US-09-336-447A-13	Sequence 13, Appl
20	81	10.4	2736	4 US-09-252-991A-30227	Sequence 30227, A
21	80	10.2	863	4 US-09-489-039A-10179	Sequence 10179, A
22	79	10.1	353	2 US-08-687-702-37	Sequence 37, Appl
23	78.5	10.0	745	4 US-09-336-115C-6	Sequence 6, Appl
24	78.5	10.0	1739	4 US-09-540-236-3739	Sequence 3739, Ap
25	78	10.0	1216	4 US-09-134-000C-5130	Sequence 5130, Ap
26	77.5	9.9	321	4 US-09-498-520A-18	Sequence 18, Appl
27	77.5	9.9	713	4 US-09-059-584-53	Sequence 53, Appl

28	77	9.8	1338	2 US-08-728-470-9	Sequence 9, Appl
29	77	9.8	1338	3 US-08-719-641-9	Sequence 9, Appl
30	77	9.8	1529	2 US-08-728-470-10	Sequence 10, Appl
31	77	9.8	1529	3 US-08-719-641-10	Sequence 10, Appl
32	77	9.8	1599	2 US-08-617-697-9	Sequence 9, Appl
33	77	9.8	1600	2 US-08-617-697-10	Sequence 10, Appl
34	77	9.8	2315	4 US-09-543-681A-5434	Sequence 5434, Ap
35	76.5	9.8	363	1 US-08-458-023B-6	Sequence 6, Appl
36	76.5	9.8	435	2 US-08-331-515A-2	Sequence 2, Appl
37	76.5	9.8	435	3 US-09-168-406A-2	Sequence 2, Appl
38	76.5	9.8	941	4 US-09-336-447A-9	Sequence 9, Appl
39	76.5	9.8	1612	1 US-08-169-927-2	Sequence 2, Appl
40	76	9.7	364	1 US-07-792-253-17	Sequence 17, Appl
41	76	9.7	1385	1 US-07-876-280-2	Sequence 2, Appl
42	76	9.7	1385	1 US-07-675-772-2	Sequence 2, Appl
43	76	9.7	1385	1 US-08-063-170-2	Sequence 2, Appl
44	76	9.7	1385	1 US-08-158-232-2	Sequence 2, Appl
45	76	9.7	1385	1 US-08-304-626-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-233-788A-59
; Sequence 59, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Kay, William W.
; APPLICANT: Colinsion, Karen S.
; APPLICANT: Clouthier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; TITLE OF INVENTION: OF SALMONELLA
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESSES:
; ADDRESSES: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,788A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 682-4900
; TELEFAX: 3723836 SEDANBERRY
; INFORMATION FOR SEQ ID NO: 59:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 151 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-233-788A-59

Query Match 87.9%; Score 687; DB 1; Length 151;
Best Local Similarity 90.1%; Pred. No. 76-60;
Matches 136; Conservative 2; Mismatches 13; Indels 0; Gaps 0;

QY 1 MLKLKVAAPALIVSGSALAGVPPWGGGAGNNGGSSGPPSTSIYQGANALNYQ 60
|||||

DB 1 MLLKVAFAAIVVSGSALAGVVPWGCGGNHNGSSGPDSTLSTIYQYGSANALALQ 60
 QY 61 LVTRVVTHEMAHAGVNGADVGAGADNSTIELTONGFRNNATIDDMNAKNSDITVGYGG 120
 DB 61 SPARKSETTITQSGNGADVGAGADNSTIELTONGFRNNATIDDMNAKNSDITVGYGG 120
 QY 121 NNAALVNQTSASDSSVVRQVGFNNATANQY 151
 DB 121 NNPALVNQTSASDSSVVRQVGFNNATANQY 151

RESULT 2

US-08-233-788A-57
 ; Sequence 57, Application US/08233788A
 ; Patent No. 5635617
 ; GENERAL INFORMATION:
 ; APPLICANT: Doran, James L.
 ; APPLICANT: Kay, William W.
 ; APPLICANT: Collinson, Karen S.
 ; APPLICANT: Clouthier, Sharon C.
 ; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
 ; TITLE OF INVENTION: OF SALMONELLA
 ; NUMBER OF SEQUENCES: 61
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Seed and Berry
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue
 ; CITY: Seattle
 ; STATE: Washington
 ; COUNTRY: U.S.A.
 ; ZIP: 98104-7092
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/233,788A
 ; FILING DATE: 26-APR-1994
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: King, Joshua
 ; REGISTRATION NUMBER: 35,570
 ; REFERENCE/DOCKET NUMBER: 920043.403C2
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (206) 622-4900
 ; TELEFAX: (206) 682-6031
 ; TELEX: 3723836 SEDANBERRY
 ; INFORMATION FOR SEQ ID NO: 57:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 120 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-233-788A-57

Query Match 64.8%; Score 507; DB 1; Length 120;
 Best Local Similarity 87.5%; Pred. No. 2,1e-42;
 Matches 98; Conservative 2; Indels 0; Gaps 0;

QY 22 VVPWGGGNNHNGSSGPDSTLSTIYQYGSANALALDQVTRVVTHEMAHAGVNGADV 81
 DB 1 VVPWGGGNNHNGSSGPDSTLSTIYQYGSANALALQSDARKSETTITQSGYGGADV 60
 QY 82 GCGADNSTIELTONGFRNNATIDDMNAKNSDITVGYGGNNAALVNQTSASD 133
 DB 61 GCGADNSTIELTONGFRNNATIDDMNAKNSDITVGYGGNNAALVNQTSASD 112

RESULT 3

US-09-056-556-204
 ; Sequence 204, Application US/09056556
 ; Patent No. 6350456
 ; GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
 APPLICANT: Skeiky, Yasir A.W.
 APPLICANT: Dillon, David C.
 TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
 NUMBER OF SEQUENCES: 241
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: SEED AND BERRY LLP
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington
 COUNTRY: USA
 ZIP: 98104-7092
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/056,556
 FILING DATE: 07-APR-1998
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: MAKI, David J.
 REGISTRATION NUMBER: 31,392
 REFERENCE/DOCKET NUMBER: 210121.457
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 622-4900
 TELEFAX: (206) 682-6031
 INFORMATION FOR SEQ ID NO: 204:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 943 amino acids
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: linear
 US-09-056-556-204

Query Match 12.8%; Score 100; DB 4; Length 943;
 Best Local Similarity 26.0%; Pred. No. 0.2;
 Matches 40; Conservative 12; Mismatches 50; Indels 50; Gaps 7;

QY 16 GSNLAGVVPWGGG-GNHN-GCGNSGPDSTLSTIYQYGSANALALDQVTRVVTHEMAHA 73
 DB 464 GSGNIGVFVWGGSGSLGNVNTIGSN-----LGIYNIQFQVNGDY-----NV 503
 QY 74 GYNGADVGAGADNSTIELTONGFRNNATIDDMNAKNSDITVGYGGN-----121
 DB 504 GFGNADPVPQFANT-----GNNTIGFANTGNNTIGIGLSGDNQCGFNINASGWS 553
 QY 122 ---NNAALVNQTSASDSSV---VROYFGNNATAN 149
 DB 554 GTNSGGLFNSGTNNVGFNAGTGNVGIANSSTGN 587

RESULT 4

US-09-072-596-199
 ; Sequence 199, Application US/09072596
 ; Patent No. 6458366
 ; GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
 APPLICANT: Skeiky, Yasir A.W.
 APPLICANT: Dillon, David C.
 APPLICANT: Campos-Neto, Antonia
 APPLICANT: Houghton, Raymond
 APPLICANT: Vedvick, Thomas S.
 APPLICANT: Twardzik, Daniel R.
 APPLICANT: Lodes, Michael J.
 APPLICANT: Hendrickson, Ronald C.
 TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
 NUMBER OF SEQUENCES: 350
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: SEED AND BERRY LLP
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle

TREAT

STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/072,596
FILING DATE: 05-MAY-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 682-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 199:
SEQUENCE CHARACTERISTICS:
LENGTH: 943 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-09-072-596-199

Query Match 12.8%; Score 100; DB 4; Length 943;
Best Local Similarity 26.0%; Pred. No. 0.2;
Matches 40; Conservative 12; Mismatches 52; Indels 50; Gaps 7;

QY 16 GSALAGVVPQMGCG-GNHN-GGNSSGPDSSTISIYOGSANAALYDQVTRVYTHMAHA 73
DB 464 GSGNIGVFNVGSGSLGNVYIGSGN-----LGIYVIGFNVGDY-----NV 503

QY 74 GYGNGADVQGADNSTIELTONGFRNNATIIDQMNKNSDITVGOYGN-----121
DB 504 GFGNAGDFVQGFANT-----GNNNIGFANTGNNNIGIGLSGDNQOGFNIAAGMNS 553

QY 122 ---NALVNTASDSSVM---VRQVFGNNATAN 149
DB 554 GTGNSGLFNSGTTNNVCIFFNAGTGNVGIANSCTGN 587

RESULT 5
US-09-477-135A-131
Sequence 131, Application US/09477135A
Patent No. 6572865
GENERAL INFORMATION:
APPLICANT: Nano, Francis
TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
FILE REFERENCE: 52888
CURRENT APPLICATION NUMBER: US/09/477,135A
CURRENT FILING DATE: 2000-01-03
PRIOR APPLICATION NUMBER: 08990823
PRIOR FILING DATE: 1997-12-15
PRIOR APPLICATION NUMBER: US 96/10375
PRIOR FILING DATE: 1996-06-14
PRIOR APPLICATION NUMBER: 60/000,254
PRIOR FILING DATE: 1995-06-15
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 131
LENGTH: 943
TYPE: PRT
ORGANISM: Mycobacterium tuberculosis
US-09-477-135A-131

Query Match 12.8%; Score 100; DB 4; Length 943;
Best Local Similarity 26.0%; Pred. No. 0.2;
Matches 40; Conservative 12; Mismatches 52; Indels 50; Gaps 7;

QY 16 GSALAGVVPQMGCG-GNHN-GGNSSGPDSSTISIYOGSANAALYDQVTRVYTHMAHA 73
DB 571 GSGNIGVFNVGSGSLGNVYIGSGN-----LGIYVIGFNVGDY-----NV 610

QY 74 GYGNGADVQGADNSTIELTONGFRNNATIIDQMNKNSDITVGOYGN-----121
DB 611 GFGNAGDFVQGFANT-----GNNNIGFANTGNNNIGIGLSGDNQOGFNIAAGMNS 660

QY 122 ---NALVNTASDSSVM---VRQVFGNNATAN 149
DB 661 GTGNSGLFNSGTTNNVCIFFNAGTGNVGIANSCTGN 694

RESULT 6
US-09-072-967-204
Sequence 204, Application US/09072967
Patent No. 6592877
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Wardzik, Daniel R.
APPLICANT: Lodes, Michael J.
APPLICANT: Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
TITLE OR INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 355
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/072,967
FILING DATE: 05-MAY-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 682-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 204:
SEQUENCE CHARACTERISTICS:
LENGTH: 943 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-09-072-967-204

Query Match 12.8%; Score 100; DB 4; Length 943;
Best Local Similarity 26.0%; Pred. No. 0.2;
Matches 40; Conservative 12; Mismatches 52; Indels 50; Gaps 7;

QY 16 GSALAGVVPQMGCG-GNHN-GGNSSGPDSSTISIYOGSANAALYDQVTRVYTHMAHA 73
DB 464 GSGNIGVFNVGSGSLGNVYIGSGN-----LGIYVIGFNVGDY-----NV 503

QY 74 GYGNGADVQGADNSTIELTONGFRNNATIIDQMNKNSDITVGOYGN-----121
DB 504 GFGNAGDFVQGFANT-----GNNNIGFANTGNNNIGIGLSGDNQOGFNIAAGMNS 553

```
QY 122 ---NALVNQTADSSVM---VRQVGFNNATAN 143
      | : | | : : | | : | |
Db 554 GTGSGIFNSGTNNVGI FNAGTGVGIANSCTGN 587
```

```

US-08-864-038A-3
RESULT 7
Sequence 3, Application US/08864038A
Patent No. 6001592
GENERAL INFORMATION:
APPLICANT: Kunio NAKASHIMA et al.
TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE cDNA, VECTOR
TITLE OF INVENTION: CONTAINING SAID cDNA, HOST CELLS TRANSFORMED WITH SAID
TITLE OF INVENTION: VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING
TITLE OF INVENTION: SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBODY
TITLE OF INVENTION: TO SAID POLYPEPTIDE
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: 812-5 Hirano
STREET: Isehinden
CITY: Tsu-city
STATE: Mie-prefecture
COUNTRY: JAPAN
ZIP: 514-01
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: Microsoft Windows 95
SOFTWARE: Word Perfect 6.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/864,038A
FILING DATE: May 28, 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 8-184459
FILING DATE: 15-July-1996
ATTORNEY/AGENT INFORMATION:
NAME: C. Bruce Hamburg
REGISTRATION NUMBER: 22,389
REFERENCE/DOCKET NUMBER: F-5610
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)986-2340
TELEFAX: (212)953-7733
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 738
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Pinctada fucata
CELL TYPE: mantle epithelial cell
FEATURE:
NAME/KEY: peptide
LOCATION: from 1 to 738
IDENTIFICATION METHOD: E (by experiment)
US-08-864-038A-3

```

[illegible]

```

RESULT 8
US-09-336-447A-5
; Sequence 5, Application US/09336447A
; Patent No. 6310190
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDERBURG, ROSS A.
; TITLE OF INVENTION: USP1 AND USP2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/09/336,447A
; CURRENT FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 892
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-336-447A-5

```

Query Match	11.9%	Score 93	DB 4	Length 892
Best Local Similarity	27.7%	Pred. No. 0.89		
Matches	41	Conservative	14	Mismatches 49
				Indels 44
				Gaps 9

QY	28	GGGNHN-----GGGNS--GPSTLSIYOGSANAALYDLVTRVTHMAHAGY	75
		::	::
Db	89	GGGDYVEAKGNVSTYGGSSNTAKGEKSTIGGDTINDANG-----TYSITIGGY	137
QY	76	-----GNGADVGGADNSTI--ELTONGFPNNATTDQMAKNSDITVGYG--GNNAL	125
		::: ::	::: ::
Db	138	YSRAIGDSSTIGGYYNATGEKSTVAGGHN-----CATGNNSTVAGGSYNQATGNNSTV	193
		::: ::	::: ::
QY	126	V-----NOTASDSVWVRQVGGNATAN	149
		::	::
Db	194	AGGSHNQATGEGSF---AAGVEINQANN	218
		::	::

```

RESULT 9
US-09-328-352-4764
/ Sequence 4764, Application US/09328352
/ Patent No. 6562958
/ GENERAL INFORMATION:
/ APPLICANT: Gary L. Breton et al.
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
/ TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
/ FILE REFERENCE: GTC99-03PA
/ CURRENT APPLICATION NUMBER: US/09/328,352
/ CURRENT FILING DATE: 1999-06-04
/ NUMBER OF SEQ ID NOS: 8252
/ SEQ ID NO 4764
/ LENGTH: 975
/ TYPE: PRF
/ ORGANISM: Acinetobacter baumannii
/ US-09-328-352-4764

```

[illegible]


```

ADDRESSER: Klauber & Jackson,
STREET: 411 Hackensack Avenue, 4th Floor
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/841,835
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/196,387
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 673 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-841-835-8

Query Match      10.5%; Score 82; DB 4; Length 673;
Best Local Similarity 29.6%; Pred. No. 7.4;
Matches 34; Conservative 14; Mismatches 59; Indels 8; Gaps 4;

QY 6 VAAFAAI-VVSGSALAGVPPQMGCGGNHNGGNSGSPDSTLSIYQGSANAALYDQ--- 61
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 99 VAAAPVPAVSTSSAGVAPNPAGSGSNSSPSSSSPTSS-SSSSPSSPGSSLAESPAA 157
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 62 -VTRVVTHEMAHAGYNGADVGGADNSTIELTONG--FRNNATIDQWNAKNSDI 113
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 158 GVSSTAPLPGGAAGPGTGVPVAVSALRELLACRNGDVSRVRLVDAANVNAKDM 212
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 14
US-09-196-387-10
Sequence 10, Application US/09196387
Patent No. 6277613
GENERAL INFORMATION:
APPLICANT: de Lange, Titia
APPLICANT: Smith, Susan
TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRP1 AND METHODS
NUMBER OF INVENTION: OF USE THEREOF
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESSES:
ADDRESSER: Klauber & Jackson
STREET: 411 Hackensack Avenue, 4th Floor
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/196,387
FILING DATE:
CLASSIFICATION:

```

```

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/095,225
FILING DATE: June 10, 1998
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 949 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-196-387-10

Query Match      10.5%; Score 82; DB 3; Length 949;
Best Local Similarity 29.6%; Pred. No. 12;
Matches 34; Conservative 14; Mismatches 59; Indels 8; Gaps 4;

QY 6 VAAFAAI-VVSGSALAGVPPQMGCGGNHNGGNSGSPDSTLSIYQGSANAALYDQ--- 61
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 99 VAAAPVPAVSTSSAGVAPNPAGSGSNSSPSSSSPTSS-SSSSPSSPGSSLAESPAA 157
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 62 -VTRVVTHEMAHAGYNGADVGGADNSTIELTONG--FRNNATIDQWNAKNSDI 113
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 158 GVSSTAPLPGGAAGPGTGVPVAVSALRELLACRNGDVSRVRLVDAANVNAKDM 212
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 15
US-09-841-835-10
Sequence 10, Application US/09841835
Patent No. 6506587
GENERAL INFORMATION:
APPLICANT: de Lange, Titia
APPLICANT: Smith, Susan
TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRP1 AND METHODS
NUMBER OF INVENTION: OF USE THEREOF
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESSES:
ADDRESSER: Klauber & Jackson
STREET: 411 Hackensack Avenue, 4th Floor
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/841,835
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/196,387
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 949 amino acids
TYPE: amino acid

```

TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-841-835-10

Query Match 10.5%; Score 82; DB 4; Length 949;
Best Local Similarity 29.6%; Pred. No. 12;
Matches 34; Conservative 14; Mismatches 59; Indels 8; Gaps 4;

Qy 6 VAAPAI-VVSGSALAGVVPOMGGGNNHGGNSGGPDSTLSIYQYGSANAALYDQ--- 61
Db 99 VAAAPVVPVPAVSTSSAAGVAPNPAGSGSNSSPSSSSPTSS-SSSSPSPGSSSLAESPEAA 157
Qy 62 -VTRVVTHEMAHAGYNGADVGQADNSTIELTQNG--FRNNATIDQNNAKNSDI 113
Db 158 GVSSTAPLGPAAAGPSTGVPAVSGALRELLACRNGDVSRYKRLVDAAVNNAKDM 212

Search completed: March 11, 2004, 18:44:53
Job time : 12.4 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:40:56 ; Search time 24.6 Seconds
(without alignments)
1296.103 Million cell updates/sec

Title: US-09-543-407-26

Perfect score: 782
Sequence: 1 NMLKRVAAFPAAIVSGSALA.....DSVMVROYFGNNATANY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 809742 seqs, 21153259 residues

Total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_AA.*

```
1: /cgn2_6/ptodata/1/pubppaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubppaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubppaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubppaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubppaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubppaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubppaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubppaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubppaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubppaa/US09C_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubppaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubppaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubppaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubppaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubppaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubppaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubppaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubppaa/US60_PUBCOMB.pep.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	116.5	14.9	445	US-10-369-493-20638	Sequence 20638, A
2	105	13.4	597	US-09-793-306-146	Sequence 146, App
3	100	12.8	943	US-09-996-634-131	Sequence 131, App
4	100	12.8	943	US-09-997-182-131	Sequence 131, App
5	100	12.8	943	US-09-997-181-131	Sequence 131, App
6	100	12.8	943	US-10-193-002-199	Sequence 199, App
7	100	12.8	943	US-10-084-843-204	Sequence 204, App
8	98.5	12.6	486	US-10-369-493-20619	Sequence 20619, A
9	93	11.9	892	US-09-952-267-5	Sequence 5, Appl
10	92	11.8	354	US-09-820-843A-21	Sequence 21, Appl
11	91.5	11.7	591	US-10-233-553-23	Sequence 23, Appl
12	91.5	11.7	894	US-10-233-553-11	Sequence 11, Appl
13	88.5	11.3	1621	US-10-185-990-10	Sequence 10, Appl
14	88.5	11.3	1626	US-10-185-990-11	Sequence 11, Appl
15	87.5	11.2	1649	US-10-369-493-18460	Sequence 18460, A

16	87	11.1	904	US-10-369-493-12420	Sequence 12420, A
17	85.5	10.9	278	US-09-810-264-28	Sequence 28, Appl
18	84	10.7	550	US-10-156-761-11286	Sequence 11286, A
19	83	10.6	400	US-10-009-823A-10	Sequence 10, Appl
20	82	10.5	174	US-10-094-749-2987	Sequence 2987, Ap
21	82	10.5	673	US-09-841-835-8	Sequence 8, Appl
22	82	10.5	949	US-09-841-835-10	Sequence 10, Appl
23	82	10.5	1017	US-10-369-493-20166	Sequence 20166, A
24	82	10.5	1327	US-09-841-835-2	Sequence 2, Appl
25	82	10.5	1327	US-09-972-115A-8	Sequence 8, Appl
26	82	10.5	1327	US-10-199-997-4	Sequence 4, Appl
27	81.5	10.4	873	US-09-952-257-13	Sequence 13, Appl
28	81	10.4	2468	US-10-246-330-4	Sequence 4, Appl
29	80.5	10.3	65	US-09-996-194-16	Sequence 16, Appl
30	80.5	10.3	1778	US-10-238-075-749	Sequence 749, App
31	80	10.2	182	US-10-238-075-1549	Sequence 1549, Ap
32	80	10.2	327	US-09-738-626-3535	Sequence 3535, Ap
33	80	10.2	406	US-10-156-761-14828	Sequence 14828, A
34	80	10.2	812	US-10-369-493-9134	Sequence 9134, Ap
35	79.5	10.2	2338	US-10-029-386-31982	Sequence 31982, A
36	79.5	10.2	2893	US-09-882-227-522	Sequence 522, App
37	79	10.1	257	US-09-880-748-1494	Sequence 1494, Ap
38	79	10.1	303	US-10-369-493-9713	Sequence 9713, Ap
39	78.5	10.0	251	US-09-880-748-1556	Sequence 1556, Ap
40	78.5	10.0	251	US-09-880-748-1840	Sequence 1840, Ap
41	78.5	10.0	535	US-10-389-566-1139	Sequence 1139, Ap
42	78.5	10.0	745	US-08-834-666A-6	Sequence 6, Appl
43	78	10.0	391	US-10-225-068-48	Sequence 48, Appl
44	78	10.0	391	US-10-225-067-102	Sequence 102, App
45	78	10.0	391	US-10-374-780A-336	Sequence 336, App

ALIGNMENTS

```
RESULT 1
US-10-369-493-20638
; Sequence 20638, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 20638
; LENGTH: 445
; TYPE: PRT
; ORGANISM: Rhodospseudomonas palustris
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(445)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-20638
```

Query Match 14.9%; Score 116.5; DB 15; Length 445;

Best Local Similarity 27.5%; Pred No. 0.0019; Gaps 41; Gaps 6;

Matches 46; Conservative 20; Mismatches 60; Indels 41; Gaps 6;

```
QY 7 AAFPA-----IVSGSALAGVVPWGSG-----NHNG-----GNSGPDSTLSIYY 50
|||
|||
|||
Db 19 AAFPAADSNVTYVINGCNDQANITQSGNGNSVGAFNGSGFLOENGLTSGA-NLITVKOS 77
|||
|||
|||
QY 51 GSNALTYDQVLYTRVYTHMAHAGYGNAGDVQAGADNSTIELTONGFRNNATITDDNNM 110
|||
|||
|||
```

Db 78 GNSNSVGRD-----10GKQSGAGNSAIF0E0GTGSDVLE0QTGTSGCAVP5GSMWNTN 129
 QY 111 -----SDITVG0YGGNNAALVNOQTASDSSVWVRQVG 141
 Db 130 DPGVENKITQDSSSSNGKSVYQDGNVNF5IKQANTGNSISVNOIG 176

RESULT 2 US-09-793-306-146

; Sequence 146, Application US/09793306
 ; Patent No. US20020098200A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Campos-Neto, Antonio
 ; APPLICANT: Skeiky, Yasir
 ; APPLICANT: Owendale, Pamela
 ; APPLICANT: Jen, Shyian
 ; APPLICANT: Lodes, Michael
 ; APPLICANT: Corixa Corporation
 ; TITLE OF INVENTION: Compounds and Methods for Diagnosis and Immunotherapy
 ; FILE REFERENCE: 014058-008740US
 ; CURRENT FILING DATE: 2001-02-26
 ; PRIOR FILING DATE: 2000-02-25
 ; PRIOR FILING DATE: 2000-02-25
 ; PRIOR APPLICATION NUMBER: US 60/185,037
 ; PRIOR APPLICATION NUMBER: US 60/223,828
 ; PRIOR FILING DATE: 2000-08-08
 ; NUMBER OF SEQ ID NOS: 164
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 146
 ; LENGTH: 597
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: mTTC#3-His
 ; US-09-793-306-146

Query Match 13.4%; Score 105; DB 9; Length 597;
 Best Local Similarity 29.0%; Pred. No. 0.04;
 Matches 40; Conservative 18; Mismatches 58; Indels 22; Gaps 7;

QY 14 VSGSALGVVPQWGGGNNHNGSSGPDSTLSIYQGSANAAALYDQLVTRVTHMAHA 73
 Db 338 LTGNLVGIGALNSGIGN-LDFGNSG--NNNIGFNSGNNNVGFNS-----GNNNF 386
 QY 74 GYGNAGADVQGGADNSTIELTONGFRNNATIDQ--WNAKNSDITVG0YGGNNAALVNOQTAS 131
 Db 387 GFGNAGDINTGFGNAGD--TWTGFGNAGFNNMIGTGNMGNEDMGVNGSGSNVGVGN--AG 442
 QY 132 DSSVWVRQYGGFNNAATAN 149
 Db 443 NOS-----VGFNGNAGTLN 455

RESULT 3 US-09-996-634-131

; Sequence 131, Application US/09996634
 ; Patent No. US20020172684A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Nano, Francis
 ; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
 ; TITLE OF INVENTION: Immunostimulatory Peptides
 ; FILE REFERENCE: 61260
 ; CURRENT APPLICATION NUMBER: US/09/996,634
 ; CURRENT FILING DATE: 2001-11-28
 ; PRIOR APPLICATION NUMBER: 09/447,135
 ; PRIOR FILING DATE: 2000-01-03
 ; PRIOR APPLICATION NUMBER: 08/990,823
 ; PRIOR FILING DATE: 1997-12-15
 ; PRIOR APPLICATION NUMBER: US 96/10375
 ; PRIOR FILING DATE: 1996-06-14
 ; PRIOR APPLICATION NUMBER: 60/000,254
 ; PRIOR FILING DATE: 1995-06-15

; NUMBER OF SEQ ID NOS: 169
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 131
 ; LENGTH: 943
 ; TYPE: PRT
 ; ORGANISM: Mycobacterium tuberculosis
 ; US-09-996-634-131

Query Match 12.8%; Score 100; DB 9; Length 943;
 Best Local Similarity 26.0%; Pred. No. 0.22;
 Matches 40; Conservative 12; Mismatches 52; Indels 50; Gaps 7;

QY 16 GSALAGVVPQWGGG--GNHN--GGGNSGPDSTLSIYQGSANAAALYDQLVTRVTHMAHA 73
 Db 571 GSGNIGVFNNVSGSLGNVNISSGN-----LGIYNIIGFNVGDY-----NV 610
 QY 74 GYGNAGADVQGGADNSTIELTONGFRNNATIDQWNAKNSDITVG0YGGN-----121
 Db 611 GFGNAGDFNQGFAANT-----GNNNIGFANTGNNNIGIGLSDNQGFNIA5GWN5 660
 QY 122 ---NAALVNOQTASDSSVW---VRQYGFNNATAN 149
 Db 661 GTGNSGLFNSGTTNNVGVIFNAGTGNVGIANSSTGN 694

RESULT 4 US-09-997-182-131

; Sequence 131, Application US/09997182
 ; Publication No. US20030049263A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Nano, Francis
 ; TITLE OF INVENTION: Mycobacterium tuberculosis DNA Sequences Encoding
 ; TITLE OF INVENTION: Immunostimulatory Peptides
 ; FILE REFERENCE: 61258
 ; CURRENT APPLICATION NUMBER: US/09/997,182
 ; CURRENT FILING DATE: 2001-11-28
 ; PRIOR APPLICATION NUMBER: 09/447,135
 ; PRIOR FILING DATE: 2000-01-03
 ; PRIOR APPLICATION NUMBER: 08/990,823
 ; PRIOR FILING DATE: 1997-12-15
 ; PRIOR APPLICATION NUMBER: US 96/10375
 ; PRIOR FILING DATE: 1996-06-14
 ; PRIOR APPLICATION NUMBER: 60/000,254
 ; PRIOR FILING DATE: 1995-06-15
 ; NUMBER OF SEQ ID NOS: 169
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 131
 ; LENGTH: 943
 ; TYPE: PRT
 ; ORGANISM: Mycobacterium tuberculosis
 ; US-09-997-182-131

Query Match 12.8%; Score 100; DB 10; Length 943;
 Best Local Similarity 26.0%; Pred. No. 0.22;
 Matches 40; Conservative 12; Mismatches 52; Indels 50; Gaps 7;

QY 16 GSALAGVVPQWGGG--GNHN--GGGNSGPDSTLSIYQGSANAAALYDQLVTRVTHMAHA 73
 Db 571 GSGNIGVFNNVSGSLGNVNISSGN-----LGIYNIIGFNVGDY-----NV 610
 QY 74 GYGNAGADVQGGADNSTIELTONGFRNNATIDQWNAKNSDITVG0YGGN-----121
 Db 611 GFGNAGDFNQGFAANT-----GNNNIGFANTGNNNIGIGLSDNQGFNIA5GWN5 660
 QY 122 ---NAALVNOQTASDSSVW---VRQYGFNNATAN 149
 Db 661 GTGNSGLFNSGTTNNVGVIFNAGTGNVGIANSSTGN 694

RESULT 5 US-09-997-181-131

; Sequence 131, Application US/09997181
 ; Publication No. US20030049269A1

GENERAL INFORMATION:
APPLICANT: Nano, Francis
TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
FILE REFERENCE: 61257
CURRENT APPLICATION NUMBER: US/09/997,181
CURRENT FILING DATE: 2001-11-28
PRIOR APPLICATION NUMBER: 09/447,135
PRIOR FILING DATE: 2000-01-03
PRIOR APPLICATION NUMBER: 08/990,823
PRIOR FILING DATE: 1997-12-15
PRIOR APPLICATION NUMBER: US 96/10375
PRIOR FILING DATE: 1996-06-14
PRIOR APPLICATION NUMBER: 60/000,254
PRIOR FILING DATE: 1995-06-15
NUMBER OF SEQ ID NOS: 169
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO: 131
LENGTH: 943
TYPE: PRT
ORGANISM: Mycobacterium tuberculosis
US-09-997-181-131

Query Match 12.8%; Score 100; DB 10; Length 943;
Best Local Similarity 26.0%; Pred. No. 0.22;
Matches 40; Conservative 12; Mismatches 52; Indels 50; Gaps 7;

QY 16 GSALAGVQWQGG-GNNH-GGNSGSPDSTISITYGSAANALYDQVTRVYTHMAHA 73
DB 571 GSGNIGVFVWGSGLGNVNGSN-----LGIYNGGNVGDY-----NV 610
QY 74 GYGNAGDVQAGADNSTIELTQNGFRNNATIDQNNANSIDITVGOYGN-----121
DB 611 GYGNAGDVQAGADNSTIELTQNGFRNNATIDQNNANSIDITVGOYGN-----121
QY 122 ---NALVNGTASDSSVM---VROYGFGNNATAN 149
DB 661 GTGNSGLFNSGTNNVGI FNA GTGNVGIANS GTGN 694

RESULT 6

US-10-193-002-199
Sequence 199, Application US/10193002
Publication No. US20030135026A1
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
Skeiky, Yasir A.W.
Dillon, Davin C.
Campos-Neto, Antonia
Houghton, Raymond
Vedvick, Thomas S.
Twardzik, Daniel R.
Lodes, Michael J.
Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
TUBERCULOSIS
NUMBER OF SEQUENCES: 350
CORRESPONDENCE ADDRESS:
ADDRESSER: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/193,002
FILING DATE: 10-Jul-2002
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/072,596
FILING DATE: 05-MAY-1998
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.41769
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 199:
SEQUENCE CHARACTERISTICS:
LENGTH: 943 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 199:
US-10-193-002-199

Query Match 12.8%; Score 100; DB 14; Length 943;
Best Local Similarity 26.0%; Pred. No. 0.22;
Matches 40; Conservative 12; Mismatches 52; Indels 50; Gaps 7;

QY 16 GSALAGVQWQGG-GNNH-GGNSGSPDSTISITYGSAANALYDQVTRVYTHMAHA 73
DB 464 GSGNIGVFVWGSGLGNVNGSN-----LGIYNGGNVGDY-----NV 503
QY 74 GYGNAGDVQAGADNSTIELTQNGFRNNATIDQNNANSIDITVGOYGN-----121
DB 504 GYGNAGDVQAGADNSTIELTQNGFRNNATIDQNNANSIDITVGOYGN-----121
QY 122 ---NALVNGTASDSSVM---VROYGFGNNATAN 149
DB 554 GTGNSGLFNSGTNNVGI FNA GTGNVGIANS GTGN 587

RESULT 7

US-10-084-843-204
Sequence 204, Application US/10084843
Publication No. US20030143243A1
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
Skeiky, Yasir A.W.
Dillon, Davin C.
Campos-Neto, Antonio
Houghton, Raymond
Vedvick, Thomas S.
Twardzik, Daniel R.
Lodes, Michael J.
Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 355
CORRESPONDENCE ADDRESS:
ADDRESSER: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/084,843
FILING DATE: 25-Feb-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/072,967
FILING DATE: 05-MAY-1998
ATTORNEY/AGENT INFORMATION:

NAME: Maki, David J.
 REGISTRATION NUMBER: 31,392
 REFERENCE/DOCKET NUMBER: 210121.411C9
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 622-4900
 TELEFAX: (206) 682-6031
 INFORMATION FOR SEQ ID NO: 204:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 943 amino acids
 TYPE: amino acid
 STRANDEDNESS: <Unknown>
 TOPOLOGY: linear
 SEQUENCE DESCRIPTION: SEQ ID NO: 204:
 US-10-084-843-204

Query Match 12.8%; Score 100; DB 14; Length 943;
 Best Local Similarity 26.0%; Pred. No. 0.22;
 Matches 40; Conservative 12; Mismatches 52; Indels 50; Gaps 7;

16 GSALAGVVPQWGG-GNH-N-GGNSGSDSTLSIYQGSANALYDQLVTRVTHMAHA 73
 464 GSGNIGVNNVSGSGISGNINISSN-----LGITNIGFQVVDY-----NV 503
 74 GYNGADVGGADNSTELTONGFRNNATIDQWAKSDITVQYGN-----121
 504 GFNAGDENQGFANT-----GNNIGFANTGNNNIGIGLSDNQGFNIASGWS 553
 122 ---NAALVNOTASDSSVM---VROVFGNNATN 149
 554 GTNSGLFNSGTTNNVGIFFNAGTGNVGIANSQTGN 587

RESULT 8
 US-10-369-493-20619
 Sequence 20619, Application US/10369493
 Publication No. US20030233675A1
 GENERAL INFORMATION:
 APPLICANT: Cao, Yongwei
 APPLICANT: Hinkle, Gregory J.
 APPLICANT: Slater, Steven C.
 APPLICANT: Goldman, Barry S.
 APPLICANT: Chen, Xianfeng
 TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 FILE REFERENCE: 38-10(52052)B
 CURRENT APPLICATION NUMBER: US/10/369,493
 PRIOR FILING DATE: 2003-02-28
 PRIOR APPLICATION NUMBER: US 60/360,039
 PRIOR FILING DATE: 2002-02-21
 NUMBER OF SEQ ID NOS: 47374
 SEQ ID NO 20619
 LENGTH: 466
 TYPE: PRT
 ORGANISM: Rhodospseudomonas palustris
 FEATURE:
 NAME/KEY: unsure
 LOCATION: (1)...(486)
 OTHER INFORMATION: unsure at all Xaa locations
 US-10-369-493-20619

Query Match 12.6%; Score 98.5; DB 15; Length 486;
 Best Local Similarity 25.7%; Pred. No. 0.14;
 Matches 37; Conservative 22; Mismatches 50; Indels 35; Gaps 6;

11 AIIVSGS-----ALAGVVPQWGGGNNHNGGNSGP-----DSTLSIYQGSANAL 57
 86 SVVAGTDYRDIVAGVPSLGGVKKAVAIIGSGSPFAPFDLSDTFPSAPFEGADGF- 144
 58 YDQLVTRVTHMAHAGYNGADVGGADNSTELTONGFRNNATIDQWAKSDITVQ- 116
 145 -----VIHTAIVAGRPFGALISGQ--NLLT-----AQSLLVAMLTADAVLGM 188
 117 -----QYGNNAALVNOTASDSSVM 136

Db 189 LPLFHTYGLGMLTLQDAGASVI 212

RESULT 9
 US-09-952-267-5
 Sequence 5, Application US/09952267
 Publication No. US20030032772A1
 GENERAL INFORMATION:
 APPLICANT: HANSEN, ERIC J.
 APPLICANT: AEBI, CHRISTOPH
 APPLICANT: COPE, LESLIE D.
 APPLICANT: MACIVER, ISOBEL
 APPLICANT: FISKE, MICHAEL J.
 APPLICANT: FREDEBERG, ROSS A.
 TITLE OF INVENTION: USP41 AND USP42 ANTIGENS OF MORAXELLA CATARRHALIS
 FILE REFERENCE: AMCY:024
 CURRENT APPLICATION NUMBER: US/09/952,267
 CURRENT FILING DATE: 2001-09-12
 PRIOR APPLICATION NUMBER: 09/336,447
 PRIOR FILING DATE: 1999-06-21
 NUMBER OF SEQ ID NOS: 98
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 5
 LENGTH: 892
 TYPE: PRT
 ORGANISM: Moraxella catarrhalis
 US-09-952-267-5

Query Match 11.9%; Score 93; DB 10; Length 892;
 Best Local Similarity 27.7%; Pred. No. 1.1;
 Matches 41; Conservative 14; Mismatches 49; Indels 44; Gaps 9;

28 GGNH-----GGNS---GPDSTLSIYQGSANALYDQLVTRVTHMAHAGY 75
 89 GGGDYNEAKNYSYVGGSSNTAKGKSTIGGDTNDANG-----TYSITGGY 137
 76 -----GNADVGGADNST--ELTONGFRNNATIDQWAKSDITVQYG---GNMAL 125
 138 YSRAGDSSITIGGIYNQATGKSTYVAGGRN---QATGNNSTYAGSGSYNATGNNSTIV 193
 126 V-----NOTASDSSVMVROVFGNNATN 149
 194 AGSHNQATGEGSF---AAGVENKANAN 218

RESULT 10
 US-09-820-843A-21
 Sequence 21, Application US/09820843A
 Publication No. US20030039963A1
 GENERAL INFORMATION:
 APPLICANT: Council of Scientific and Industrial Research
 TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTE
 FILE REFERENCE: Q63915
 CURRENT APPLICATION NUMBER: US/09/820,843A
 CURRENT FILING DATE: 2001-03-30
 NUMBER OF SEQ ID NOS: 118
 SOFTWARE: PatentIn version 3.0
 SEQ ID NO 21
 LENGTH: 354
 TYPE: PRT
 ORGANISM: M. tuberculosis
 FEATURE:
 NAME/KEY: misc.feature
 OTHER INFORMATION: PPE
 NAME/KEY: misc.feature
 OTHER INFORMATION: g11781260
 US-09-820-843A-21

Query Match 11.8%; Score 92; DB 10; Length 354;
 Best Local Similarity 27.6%; Pred. No. 0.42;
 Matches 34; Conservative 12; Mismatches 43; Indels 34; Gaps 5;

QY 27 GGGGNNHGGSSGPDSTLSIYOGSANAALYDQVTVRVVTHEMAHAGYNGADVGQAD 86
 DB 46 GNNGNFPGSGYNG-SNNIGFNGTSGN-----FGFGN-----TGN 81
 QY 87 NSTIETLONFRNNATIDOWNAKNSDITVGOYGNNALVNOTASDSVWVROVFGNNA 146
 DB 82 NIGIGLITGDC---QIGIGLINSNGNIGFNSGTGNVGLFNSGTGN-----VGRNSG 131
 QY 147 TAN 149
 DB 132 TAN 134

RESULT 11
 US-10-233-553-23
 ; Sequence 23, Application US/10233553
 ; Publication No. US20030125285A1
 ; GENERAL INFORMATION:
 ; APPLICANT: NIPPON SHINYAKU CO., LTD.
 ; APPLICANT: HIRABAYASHI, Kazuko
 ; APPLICANT: YANO, Junichi
 ; TITLE OF INVENTION: Method of examining the efficacy of therapy with nucleic acid
 ; FILE REFERENCE: B-345
 ; CURRENT APPLICATION NUMBER: US/10/233,553
 ; CURRENT FILING DATE: 2002-12-19
 ; PRIOR APPLICATION NUMBER: JP 2001-267385
 ; PRIOR FILING DATE: 2001-09-04
 ; NUMBER OF SEQ ID NOS: 31
 ; SEQ ID NO 23
 ; LENGTH: 591
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-233-553-23

Query Match 11.7%; Score 91.5; DB 14; Length 591;
 Best Local Similarity 27.1%; Pred. No. 0.89;
 Matches 35; Conservative 15; Mismatches 32; Indels 47; Gaps 7;

QY 27 GGGG-----NHNCGGNSGPDSTLSIYOGSANAALYDQVTVRVVTHEMAHAGYGN 77
 DB 492 GGGGSDYNTYKRNYSKNGSGSGSGNS-----YSGGAST-----NPGSHGTYGG 535
 QY 78 GADVQGGADNSTIETLONFRNNATIDOWNAKNSDITVGOYGNNALVNOTASDSVWV 137
 DB 536 GSGCG-----SSYQKGQGYGYSQ-----NYNSPGS-----CG-----NYSGPSSYSQ 573
 QY 138 ROYFGNNA 146
 DB 574 SOGGYGRNA 582

RESULT 12
 US-10-233-553-11
 ; Sequence 11, Application US/10233553
 ; Publication No. US20030125285A1
 ; GENERAL INFORMATION:
 ; APPLICANT: NIPPON SHINYAKU CO., LTD.
 ; APPLICANT: HIRABAYASHI, Kazuko
 ; APPLICANT: YANO, Junichi
 ; TITLE OF INVENTION: Method of examining the efficacy of therapy with nucleic acid
 ; FILE REFERENCE: B-345
 ; CURRENT APPLICATION NUMBER: US/10/233,553
 ; CURRENT FILING DATE: 2002-12-19
 ; PRIOR APPLICATION NUMBER: JP 2001-267385
 ; PRIOR FILING DATE: 2001-09-04
 ; NUMBER OF SEQ ID NOS: 31
 ; SEQ ID NO 11
 ; LENGTH: 894
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-233-553-11

Query Match 11.7%; Score 91.5; DB 14; Length 894;
 Best Local Similarity 27.1%; Pred. No. 1.5;
 Matches 35; Conservative 15; Mismatches 32; Indels 47; Gaps 7;
 QY 27 GGGG-----NHNCGGNSGPDSTLSIYOGSANAALYDQVTVRVVTHEMAHAGYGN 77
 DB 795 GGGGSDYNTYKRNYSKNGSGSGSGNS-----YSGGAST-----NPGSHGTYGG 838
 QY 78 GADVQGGADNSTIETLONFRNNATIDOWNAKNSDITVGOYGNNALVNOTASDSVWV 137
 DB 839 GSGCG-----SSYQKGQGYGYSQ-----NYNSPGS-----CG-----NYSGPSSYSQ 876
 QY 138 ROYFGNNA 146
 DB 877 SOGGYGRNA 885

RESULT 13
 US-10-185-990-10
 ; Sequence 10, Application US/10185990
 ; Publication No. US20030073109A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Pan, Jae-Gu
 ; TITLE OF INVENTION: JAE GU PAN ET AL
 ; FILE REFERENCE: 02589,000100
 ; CURRENT APPLICATION NUMBER: US/10/185,990
 ; CURRENT FILING DATE: 2002-06-28
 ; NUMBER OF SEQ ID NOS: 11
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 10
 ; LENGTH: 1621
 ; TYPE: PRT
 ; ORGANISM: BACILLUS SUBTILIS
 US-10-185-990-10

Query Match 11.3%; Score 88.5; DB 14; Length 1621;
 Best Local Similarity 27.1%; Pred. No. 6.3;
 Matches 42; Conservative 21; Mismatches 41; Indels 51; Gaps 11;

QY 16 GSALG-----VVPWCG-----GGNHN-----GGNSGSPDSTLSIYOGSA 53
 DB 183 GSTLSGDHNSRLIAGYSNETHGNHSDLIAGYSTGTAGYSTGYSGEDSSLTA-CYGST 241
 QY 54 NAALYDQVTVRVVTHEMAHAGYNGADVGQADNSTIETLONFRNNATIDOWNAKNSDI 113
 DB 242 QTAQSGSNLT-----AGYSTGCTA--GSDSSLTA-----CYGSTGT-----SGGDSST 282
 QY 114 TVGOYGNNALVNOTASDSVWVROVFGNNA 148
 DB 283 TAG-YGS-----TQTAQSGSNLT--AGYSTGCTA 308

RESULT 14
 US-10-185-990-11
 ; Sequence 11, Application US/10185990
 ; Publication No. US20030073109A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Pan, Jae-Gu
 ; TITLE OF INVENTION: JAE GU PAN ET AL
 ; FILE REFERENCE: 02589,000100
 ; CURRENT APPLICATION NUMBER: US/10/185,990
 ; CURRENT FILING DATE: 2002-06-28
 ; NUMBER OF SEQ ID NOS: 11
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 11
 ; LENGTH: 1626
 ; TYPE: PRT
 ; ORGANISM: Bacillus subtilis
 US-10-185-990-11

Query Match 11.3%; Score 88.5; DB 14; Length 1626;
 Best Local Similarity 27.1%; Pred. No. 6.3;
 Matches 42; Conservative 21; Mismatches 41; Indels 51; Gaps 11;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:24:14 ; Search time 10.3 Seconds
(without alignments)
1410.186 Million cell updates/sec

Title: US-09-543-407-26

Perfect score: 782

Sequence: 1 MKLLKVAAPFAIVSGSALA.....DSSVMRVQVGFNNATATNOY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 78:*

1: p1r1:*

2: p1r2:*

3: p1r3:*

4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	692	88.5	151	2 JC6039	fimbria protein ag
2	692	88.5	151	2 A10635	major curlin chain
3	523	66.9	151	2 S70788	curlin protein csg
4	501.5	64.1	152	2 D90806	curlin major subun
5	501.5	64.1	152	2 H85665	hypothetical prote
6	114.5	14.6	145	2 AD3143	conserved hypochet
7	114.5	14.6	145	2 H89144	hypothetical prote
8	113	14.5	2174	2 E95965	hypothetical glyci
9	112	14.3	151	2 S70787	curlin nucleator p
10	112	14.3	151	2 C90806	minor curlin subun
11	112	14.3	151	2 G85665	curlin minor chain
12	107.5	13.7	1322	2 S07053	ice nucleation pro
13	107	13.7	151	2 JC6040	fimbria protein ag
14	107	13.7	151	2 AH0635	nucleation compone
15	106	13.6	1034	2 JC2143	ice nucleation act
16	105.5	13.5	645	2 F70825	probable PPR prote
17	105	13.4	590	2 E70946	probable PPR prote
18	104.5	13.4	1258	2 D70604	ice nucleation pro
19	102.5	13.1	552	2 D70604	probable PPR prote
20	102.5	13.1	1567	2 S11672	ice nucleation pro
21	102.5	13.1	1565	2 E97835	hypothetical prote
22	100	12.8	3300	2 D70575	probable PPR prote
23	97.5	12.5	1028	2 A56038	DNA-binding protei
24	97.5	12.5	1213	2 S16356	ovo protein - fru
25	97	12.4	615	2 E70663	probable PPR prote
26	96.5	12.3	447	2 G84687	probable disease r
27	95.5	12.2	1651	2 JC1340	outer membrane pro
28	95	12.1	652	2 E97857	cell surface anti
29	94	12.0	599	2 B42049	leishmanolysin (EC

30	94	12.0	599	2 A44951	leishmanolysin (EC
31	93	11.9	434	2 E70768	hypothetical glyci
32	92.5	11.8	575	2 S35327	protein kinase csg
33	92.5	11.8	639	2 C42049	leishmanolysin (EC
34	92.5	11.8	963	2 B70524	probable PPR prote
35	92.5	11.8	1053	2 B70987	probable PPR prote
36	92	11.8	354	2 B70663	probable PPR prote
37	92	11.8	1436	2 B70520	probable PPR prote
38	92	11.8	1748	2 S42136	cnjB protein - Tet
39	91.5	11.7	582	2 F70675	probable PPR prote
40	91	11.6	678	2 A70762	probable PPR prote
41	90.5	11.6	1200	1 SNPSO	ice nucleation pro
42	90.5	11.6	3716	2 E70969	probable PPR prote
43	90	11.5	586	2 T26667	hypothetical prote
44	90	11.5	1390	2 T14004	trfA protein - sli
45	89.5	11.4	1317	2 A54831	nuclear pore compl

ALIGNMENTS

RESULT 1

JC6039 fimbria protein agfa precursor - Salmonella enteritidis

C/Species: Salmonella enteritidis

C/Date: 31-Dec-1996 #sequence, revision 31-Dec-1996 #ext_change 08-Oct-1999

C/Accession: JC6039; PC6015; A44898

R/Collinson, S.K.; Cloughier, S.C.; Doran, J.L.; Baner, P.A.; Kay, W.W.

A/Title: Salmonella enteritidis agfBAC operon encoding thin, aggregative fimbriae.

A/Reference number: JC6039; PMID:9614512; PMID:8550497

A/Accession: JC6039

A/Molecule type: DNA

A/Residues: 1-151 <COL>

A/Cross-references: GB:U43280; NID:G1184712; PIDN:AAC43599.1; PID:G1184714

A/Accession: PC6015

A/Molecule type: protein

A/Residues: 21-52 <CO2>

A/Experimental source: strain 27655-3b

R/Collinson, S.K.; Emdy, L.; Muller, K.H.; Trust, T.J.; Kay, W.W.

J. Bacteriol. 173, 4773-4781, 1991

A/Title: Purification and characterization of thin, aggregative fimbriae from Salmonella

A/Reference number: A44898; PMID:91310586; PMID:1677357

A/Accession: 27655

A/Status: preliminary

A/Molecule type: protein

A/Residues: 21-33 <CO3>

C/Note: sequence extracted from NCBI backbone (NCBIP:45936)

C/Genetic:

A/Gene: agfa

C/Function:

A/Description: major component of thin aggregative fimbriae

A/Note: fimbriae bind to fibronectin, plasminogen, tissue plasminogen activator

C/Keywords: fimbria

F/1-20/Domain: signal sequence #status predicted <Sig>

F/21-151/Product: fimbria protein agfa #status experimental <Mat>

Query Match 88.5%; Score 692; DB 2; Length 151;
Best Local Similarity 90.7%; Pred. No. 3, 6e-51;
Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

QY	1	MKLLKVAAPFAIVSGSALAGVVPVGGGNNHGGSSGPDSTLSIYGYGSAANALYDQ	60
DB	1	MKLLKVAAPFAIVSGSALAGVVPVGGGNNHGGSSGPDSTLSIYGYGSAANALYDQ	60
QY	61	LVTYVTHMAAGYNGADVGGADNSTIELTONGFRNNATIDWNAKNSDITVQYGG	120
DB	61	SDARKSETITGSGYNGADVGGADNSTIELTONGFRNNATIDWNAKNSDITVQYGG	120
QY	121	NNAAALVNTASDSSVMRVQVGFNNATATNOY	151
DB	121	NNAAALVNTASDSSVMRVQVGFNNATATNOY	151

Db 121 NNAALVNOTASDSSVWVRQVGFNNATANOY 151

RESULT 2

major curlin chain precursor [imported] - Salmonella enterica subsp. enterica serovar Typhimurium

C:Species: Salmonella enterica subsp. enterica serovar Typhimurium

A:Note: this species has also been called Salmonella typhimurium

C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002

C:Accession: A10635

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.

A:Authors: Park, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhimurium

A:Reference number: AB0502; MUID:21534947; PMID:11677608

A:Accession: A10635

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-151 <PAR>

A:Cross-references: GB:AL513382; PIND:CAD08268.1; PID:g16502315; GSPDB:GN00176

C:Genetics:

A:Gene: STY1181

Query Match

Best Local Similarity 88.5%; Score 692; DB 2; Length 151;

Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

Db 1 MLLKVAAPAAIVVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYOGSANAALQ 60

1 MLLKVAAPAAIVVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYOGSANAALQ 60

QY 61 LVTRVVTHEMAHAGYGAGDVGAGADNSTIELTONGFRNNATIDQNAKNSDITVGOYG 120

61 SPARSETTTTGGGNGADVGAGADNSTIELTONGFRNNATIDQNAKNSDITVGOYG 120

Db 121 NNAALVNOTASDSSVWVRQVGFNNATANOY 151

121 NNAALVNOTASDSSVWVRQVGFNNATANOY 151

QY 121 NNAALVNOTASDSSVWVRQVGFNNATANOY 151

121 NNAALVNOTASDSSVWVRQVGFNNATANOY 151

Db 121 NNAALVNOTASDSSVWVRQVGFNNATANOY 151

121 NNAALVNOTASDSSVWVRQVGFNNATANOY 151

RESULT 3

curlin protein csgA precursor - Escherichia coli (strain K-12)

N:Alternate names: csgA protein; major curlin protein

C:Species: Escherichia coli

C:Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 01-Mar-2002

C:Accession: S70788; G64846; S31202; S34560; S34559

R:Hammar, M.; Arngvist, A.; Bian, Z.; Olsen, A.; Normark, S.

Mol. Microbiol. 18, 661-670, 1995

A:Title: Expression of two csg operons is required for production of fibronectin- and Cc

A:Reference number: S70783; MUID:96414468; PMID:8817489

A:Accession: S70788

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-151 <HAM>

A:Cross-references: EMBL:X90754; NID:g1147558; PIND:CAA62282.1; PID:g1147564

A:Experimental source: strain K12, substrain W3110

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co

A:Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: G64846

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-151 <BLAT>

A:Cross-references: GB:AE000205; GB:U00096; NID:g1787265; PIND:AACT4126.1; PID:g1787279;

A:Experimental source: strain K-12, substrain MG1655

R:Olsen, A.; Arngvist, A.; Hammar, M.; Sukupolvi, S.; Normark, S.

Mol. Microbiol. 7, 523-536, 1993

A:Title: The Rpos sigma factor relieves H-NS-mediated transcriptional repression of csg

A:Reference number: S31202; MUID:93211294; PMID:8459772

A:Accession: S31202

A:Molecule type: DNA

A:Residues: 1-6, 'V', '8-151 <OLS>

A:Cross-references: EMBL:L04979

A:Accession: S34550

A:Molecule type: protein

A:Residues: 21-42; 44-50 <OLS>

R:Olsen, A.N.; Arngvist, A.M.

A:Note: submitted to the EMBL Data Library, October 1992

A:Reference number: S34559

A:Accession: S34559

A:Molecule type: DNA

A:Residues: 1-133, 'RORDSGWLV' <OLS>

A:Cross-references: EMBL:L04979; NID:g290424; PIND:AAA23616.1; PID:g290425

A:Experimental source: strain K-12, substrain W3110

C:Genetics:

A:Gene: csgA

A:Map position: 23.15

C:Function:

A:Description: major component of wild-type curli; interaction between CsgA and CsgB tr

A:Note: curli are thin, coiled fibers expressed on the surface of Escherichia coli that

F:1-20/Domain: signal sequence #status predicted <SIG>

F:21-151/Product: curlin #status experimental <MAT>

Query Match

Best Local Similarity 66.3%; Score 523; DB 2; Length 151;

Matches 104; Conservative 18; Mismatches 29; Indels 0; Gaps 0;

Db 1 MLLKVAAPAAIVVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYOGSANAALQ 60

1 MLLKVAAPAAIVVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYOGSANAALQ 60

QY 61 LVTRVVTHEMAHAGYGAGDVGAGADNSTIELTONGFRNNATIDQNAKNSDITVGOYG 120

61 TDARNSDLRTTQGGNGADVGAGDSSDIDLQRFNGSATIDQNGNNSDITVGOYG 120

Db 121 NNAALVNOTASDSSVWVRQVGFNNATANOY 151

121 NNAALVNOTASDSSVWVRQVGFNNATANOY 151

QY 121 NNAALVNOTASDSSVWVRQVGFNNATANOY 151

121 NNAALVNOTASDSSVWVRQVGFNNATANOY 151

Db 121 NNAALVNOTASDSSVWVRQVGFNNATANOY 151

121 NNAALVNOTASDSSVWVRQVGFNNATANOY 151

RESULT 4

curlin major subunit CsgA [imported] - Escherichia coli (strain O157:H7, substrain RIMD

C:Species: Escherichia coli

C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001

C:Accession: D90806

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G

gasawara, N.; Yasunaga, T.; Kihara, S.; Shiba, T.; Hattori, M.; Shingawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and ge

A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: D90806

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-152 <HAY>

A:Cross-references: GB:BA00007; PIND:BA834843.1; PID:g13360880; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RIMD 0509952

C:Genetics:

A:Gene: Ecg1420

Query Match

Best Local Similarity 64.1%; Score 501.5; DB 2; Length 152;

Matches 102; Conservative 19; Mismatches 30; Indels 1; Gaps 1;

Db 1 MLLKVAAPAAIVVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYOGSANAALQ 59

1 MLLKVAAPAAIVVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYOGSANAALQ 59

QY 60 QLVTRVVTHEMAHAGYGAGDVGAGADNSTIELTONGFRNNATIDQNAKNSDITVGOYG 119

Db 61 QADNRNSDLTITTHGGGNGADVGGGSDSSIDLTRQGFNGSATLDDQNGKDSHTVYQFG 120
 QY 120 GNNAAALVNQTAASDSSVWVRQVGFNNATANOY 151
 Db 121 GNGAANDQTAASNTVWVTVQVGFNNATANOY 152

RESULT 5

H85665
 hypothetical protein csgA [imported] - Escherichia coli (strain O157:H7, substrain EDL93
 C/Species: Escherichia coli
 C/Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 C/Accession: H85665
 R/Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
 Miller, L.; Grochbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
 Nature 409, 529-533, 2001
 A/Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A/Reference number: A85480; MUID:21074935; PMID:11206551
 A/Accession: H85665
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-152 <STO>
 A/Cross-references: GB:AE005174; NID:g12514574; PIDN:AA655788.1; GSPDB:GN00145; UMGF:216
 A/Experimental source: strain O157:H7, substrain EDL933
 C/Genetics:
 A/Map position: csgA

Query Match 64.1%; Score 501.5; DB 2; Length 152;
 Best Local Similarity 67.1%; Pred. No. 3,4e-35;
 Matches 102; Conservative 19; Mismatches 30; Indels 1; Gaps 1;
 QY 1 MKLLKVAAPAIIVSGSALAGVPPQW-GGGNNHNGGSSGPDSTLSIYQGSANALYD 59
 Db 1 MKLLKVAAPAIIVSGSALAGVPPQWGGGNNHNGGSSGPDSTLSIYQGSANALYD 60
 QY 60 QLVTRVYTHMAHAGYNGADVGGGADNSTIELTONGFPNNATTIDQNAKNSDITV 119
 Db 61 QADNRNSDLTITTHGGGNGADVGGGSDSSIDLTRQGFNGSATLDDQNGKDSHTVYQFG 120
 QY 120 GNNAAALVNQTAASDSSVWVRQVGFNNATANOY 151
 Db 121 GNGAANDQTAASNTVWVTVQVGFNNATANOY 152

RESULT 6

AD3143
 conserved hypothetical protein Atu4768 [imported] - Agrobacterium tumefaciens (strain C5
 C/Species: Agrobacterium tumefaciens
 C/Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
 C/Accession: AD3143
 R/Wood, D.W.; Seubert, J.C.; Kaul, R.; Monke, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.
 erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McCell
 Science 294, 2317-2323, 2001
 A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
 ster, E.W.
 A/Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
 A/Reference number: AB2577; MUID:21608550; PMID:11743193
 A/Accession: AD3143
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-145 <KUR>
 A/Cross-references: GB:AE008689; PIDN:AL45562.1; PID:g17743277; GSPDB:GN00187
 A/Experimental source: strain C58 (Dupont)
 C/Genetics:
 A/Map position: linear chromosome

Query Match 14.6%; Score 114.5; DB 2; Length 145;
 Best Local Similarity 25.2%; Pred. No. 0.0088;
 Matches 39; Conservative 26; Mismatches 59; Indels 31; Gaps 4;

QY 3 LKVAAPAIIVSGSALAGVPPQW-----GNNHNGGSSGPDSTLSIYQGSANA 55
 Db 1 MIRKSFIAASALVALVGLISAAPAMANDVRIEQYGMNSAGAGQEGYGNRIRTYONGGYN- 59
 QY 56 ALVDQLVTRVYTHMAHAGYNGADVGGGADNSTIELTONGFPNNATTIDQNAKNSDITV 115
 Db 60 -----RIVGQY---GRHNLSAVGQBGHDYGSSTTONGNNAVGI----- 96

RESULT 7

H98144
 hypothetical protein AGR_L_228 [imported] - Agrobacterium tumefaciens (strain C58, Cere
 C/Species: Agrobacterium tumefaciens
 C/Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
 C/Accession: H98144
 R/Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman
 A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappae, C.; Markelz, B.
 Science 294, 2323-2328, 2001
 A/Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tu
 A/Reference number: A97359; MUID:21608551; PMID:11743194
 A/Accession: H98144
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-145 <KUR>
 A/Cross-references: GB:AE007870; PIDN:AAK88682.1; PID:g15158413; GSPDB:GN00170
 C/Genetics:
 A/Map position: linear chromosome

Query Match 14.6%; Score 114.5; DB 2; Length 145;
 Best Local Similarity 25.2%; Pred. No. 0.0088;
 Matches 39; Conservative 26; Mismatches 59; Indels 31; Gaps 4;
 QY 3 LKVAAPAIIVSGSALAGVPPQW-----GNNHNGGSSGPDSTLSIYQGSANA 55
 Db 1 MIRKSFIAASALVALVGLISAAPAMANDVRIEQYGMNSAGAGQEGYGNRIRTYONGGYN- 59
 QY 56 ALVDQLVTRVYTHMAHAGYNGADVGGGADNSTIELTONGFPNNATTIDQNAKNSDITV 115
 Db 60 -----RIVGQY---GRHNLSAVGQBGHDYGSSTTONGNNAVGI----- 96
 QY 116 GQYGNNAALVNQTAASDSSVWVRQVGFNNATANO 150
 Db 97 GQFGSHHTITLTDGNGNTAAGVQVGRGCSANVSQ 131

RESULT 8

E95965
 hypothetical glycine-rich protein [imported] - Sinorhizobium meliloti (strain 1021) mag
 C/Species: Sinorhizobium meliloti
 C/Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
 C/Accession: E95965
 R/Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Herna
 Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
 A/Title: The complete sequence of the 1.683-kb pSymB megaplasmid from the N2-fixing end
 A/Reference number: A95842; MUID:21396508; PMID:11481431
 A/Accession: E95965
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-2174 <KUR>
 A/Cross-references: GB:AL591985; PIDN:CA49389.1; PID:g15140875; GSPDB:GN00167
 A/Experimental source: strain 1021, megaplasmid pSymB
 R/Galibert, F.; Finan, T.M.; Long, S.R.; Punler, A.; Abola, P.; Ampe, F.; Barloy-Hubler
 pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federapfel, N.A.; Fisher, R.F.
 L.; Hyman, R.W.; Jones, T.
 Science 293, 668-672, 2001
 A/Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kise, B.; Komp, C.; Lelaure
 hebaull, P.; Vandemol, M.; Vorholter, F.J.; Weidner, S.; Wallis, D.H.; Wong, K.; Yeh, K
 A/Title: The composite genome of the legume symbiont Sinorhizobium meliloti.

A:Reference number: A96039; MUID:21368234; PMID:11474104
 A:Contents: annotation
 C:Genetics:
 A:Gene: SMD21548
 A:Genome: plasmid

Query Match 14.3%; Score 113; DB 2; Length 2174;
 Best Local Similarity 27.0%; Pred. No. 0.23;

Matches 40; Conservative 21; Mismatches 51; Indels 36; Gaps 7;

QY 11 AIVSGSALAGVFPQ--WCGGNNHNGSGPDTLSIYQYGSANALYDQVTVRVVTH 68
 DB 693 AATATAGAGVILAGISGGGNN---CGNATGDAFGSFGICGGGCG----- 737
 QY 69 EWAHAGYNGADV-----QCADNSTI--ELTNGFRNNTIDQWNAK--NSDITV 115
 DB 738 ---GGYANTAVNPFKGLTTLTQGSHPAGIVAOSVGGGCGTGTSSAGIGFTAVAV 793
 QY 116 GQYGGNNA--LVNOTASDSVWVRQVG 141
 DB 794 GGTGNGAGGEVSUTSDAIRTQGG 821

RESULT 9

750787
 curlin nucleator protein csb precursor - Escherichia coli (strain K-12)

C:Species: Escherichia coli
 C>Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 01-Mar-2002

C/Accession: S70787; F64846
 R/Hammar, M.; Arngvist, A.; Bian, Z.; Olsen, A.; Normark, S.
 Mol. Microbiol. 18, 661-670, 1995

A>Title: Expression of two csb operons is required for production of fibronectin- and C

A:Reference number: S70783; MUID:96414468; PMID:8817489

A:Accession: S70787
 A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-151 <HAM>
 A:Cross-references: EMBL:X90754; NID:g1147559; PIDD:CA62281.1; PID:g1147563

A:Experimental source: strain K12, substrain W3110
 A>Note: The nucleotide sequence was submitted to the EMBL Data Library, August 1995

R/Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
 .A.; Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997

A>Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: F64846
 A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-151 <BLAT>
 A:Cross-references: GB:AE000205; GB:U00096; NID:g1787265; PIDD:AAC74125.1; PID:g1787278;

A:Experimental source: strain K-12, substrain MG1655

A:Genetics:

A:Gene: csbB

A:Map position: 23.15

C:Function:

A:Description: minor component of wild-type curl; interaction between CsgA and CsgB tri
 A:Note: curl are thin, coiled fibers expressed on the surface of Escherichia coli that
 and H-kinnogen; in the absence of CsgA, CsgB can self-assemble into polymers
 F1-21/Dominant: signal sequence #status predicted <SIG>
 F122-151/Product: minor curlin chain #status predicted <MAN>

Query Match 14.3%; Score 112; DB 2; Length 151;
 Best Local Similarity 29.8%; Pred. No. 0.015;

Matches 34; Conservative 17; Mismatches 49; Indels 14; Gaps 3;

QY 51 GSANALYDQVLT--RVVTHEMAHAGYNGADVGGADNSTLTETONGFR----- 98
 DB 17 GIAAAGYDLANSEYFAVNELSKSSFNQAAITIGAGTNNNSAQLRGSGSKLAVAAQEGS 76
 QY 99 -NNATIDQWNAKNSDITVQYGGNNAALVNQASDSVWVRQVGRNNATANQY 151
 DB 77 SNRAKIDQTDYNL-AYIDQASANDASISQAGYGTAMTIQSGSKNKANTQY 129

RESULT 10

minor curlin subunit precursor CsgB [imported] - Escherichia coli (strain O157:H7, sub

C:Species: Escherichia coli
 C/Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001

C/Accession: C90806
 R/Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G

gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 8, 11-22, 2001

A>Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen

A:Reference number: A9629; MUID:2136231; PMID:11258796

A:Accession: C90806

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-151 <HAV>
 A:Cross-references: GB:BA000007; PIDD:BAH34842.1; PID:g13360879; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RIMD 0509952

A:Genetics:

A:Gene: EcsB1419

Query Match 14.3%; Score 112; DB 2; Length 151;
 Best Local Similarity 29.8%; Pred. No. 0.015;

Matches 34; Conservative 17; Mismatches 49; Indels 14; Gaps 3;

QY 51 GSANALYDQVLT--RVVTHEMAHAGYNGADVGGADNSTLTETONGFR----- 98
 DB 17 GIAAAGYDLANSEYFAVNELSKSSFNQAAITIGAGTNNNSAQLRGSGSKLAVAAQEGS 76
 QY 99 -NNATIDQWNAKNSDITVQYGGNNAALVNQASDSVWVRQVGRNNATANQY 151
 DB 77 SNRAKIDQTDYNL-AYIDQASANDASISQAGYGTAMTIQSGSKNKANTQY 129

RESULT 11

curlin minor chain precursor, CsgA homolog [imported] - Escherichia coli (strain O157:H

C:Species: Escherichia coli
 C/Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001

C/Accession: G85665
 R/Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayh

iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca
 Nature 409, 529-533, 2001

A>Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: G85665

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-151 <STO>
 A:Cross-references: GB:AE005174; NID:g12514573; PIDD:AAQ55787.1; GSPDB:GN00145; UWGP:Z1

A:Experimental source: strain O157:H7, substrain EDL933

A:Genetics:

A:Gene: csbB

Query Match 14.3%; Score 112; DB 2; Length 151;
 Best Local Similarity 29.8%; Pred. No. 0.015;

Matches 34; Conservative 17; Mismatches 49; Indels 14; Gaps 3;

QY 51 GSANALYDQVLT--RVVTHEMAHAGYNGADVGGADNSTLTETONGFR----- 98
 DB 17 GIAAAGYDLANSEYFAVNELSKSSFNQAAITIGAGTNNNSAQLRGSGSKLAVAAQEGS 76
 QY 99 -NNATIDQWNAKNSDITVQYGGNNAALVNQASDSVWVRQVGRNNATANQY 151
 DB 77 SNRAKIDQTDYNL-AYIDQASANDASISQAGYGTAMTIQSGSKNKANTQY 129

RESULT 12

ice nucleation protein inaA - Erwinia ananas

C:Species: Erwinia ananas
 C/Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 26-Aug-1999

C/Accession: S07053
 R/Abe, K.; Watabe, S.; Emori, Y.; Watanabe, M.; Arai, S.
 FEBS Lett. 258, 297-300, 1989
 A/Title: An ice nucleation active gene of *Erwinia ananas*. Sequence similarity to those of
 A/Reference number: S07053; MUID:90092494; PMID:2599095
 A/Accession: S07053
 A/Status: not compared with conceptual translation
 A/Molecule type: DNA
 A/Residues: 1-1322 <ABE>
 A/Cross-references: GB:X17316; NID:g296095; PIDN:CA35194.1; PID:g296096
 C/Superfamily: ice nucleation protein

Query Match 13.7%; Score 107.5; DB 2; Length 1322;
 Best Local Similarity 29.9%; Pred. No. 0.38;
 Matches 38; Conservative 19; Mismatches 21; Indels 49; Gaps 8;

QY 34 GGGNSGPDSTLSIYQGSANALYDQVTRVVTHEMAHAGYNGADVQ-GADNST-----IELTNGF 93
 DB 933 GSTSTAGPDSLL-IAGGSGTQTAGYNSILT-----AGYGS-----T 967
 QY 94 QNGFRNATIDQWNAKNSDITVQYQ-----GNNALV-----NOTASDSSVMVRYQYFG 143
 DB 968 QTG-----QENSDLTG-TGSGSTAGYESSLIAGYGTQTASFKSTLM--AGYG 1013
 QY 144 NNATANO 150
 DB 1014 SSQTARE 1020

RESULT 13

JC6040
 fimbriin protein agfB precursor - *Salmonella enteritidis*
 C/Species: *Salmonella enteritidis*
 C/Date: 31-Dec-1996 #sequence_revision 31-Dec-1996 #text_change 08-Oct-1999
 C/Accession: JC6040
 R/Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Baner, P.A.; Kay, W.W.
 J. Bacteriol. 178, 662-667, 1996
 A/Title: *Salmonella enteritidis* agfBAC operon encoding thin, aggregative fimbriae.
 A/Reference number: JC6039; MUID:96146512; PMID:8550497
 A/Accession: JC6040
 A/Molecule type: DNA
 A/Residues: 1151 <COL>
 A/Cross-references: GB:U43280; NID:g1184712; PIDN:AAC43598.1; PID:g1184713
 A/Experimental source: strain 276755-3b
 C/Genetic8:
 A/Gene: agfB
 C/Function:
 A/Description: minor component of thin aggregative fimbriae
 A/Note: fimbriae bind to fibronectin, plasminogen, tissue plasminogen activator
 C/Keywords: fimbria
 F/1-21/Domain: signal sequence #status predicted <SIG>
 F/22-151/Product: fimbriin protein agfB #status predicted <MAT>

Query Match 13.7%; Score 107; DB 2; Length 151;
 Best Local Similarity 30.4%; Pred. No. 0.039;
 Matches 35; Conservative 17; Mismatches 47; Indels 16; Gaps 5;

QY 51 GSANALYD--QLVTRVTHEMAHAGYNGADVQ-GADNST-----IELTNGF 97
 DB 17 GATATNTYDLARSEYNFAVELSKSSFNQAIIIGQVGTDSARVROGSKLSVTSQEG 76
 QY 98 RNNATIDQWNAKNSDITVQYQGSANALVNOTASDSSVMVRYQYFGNNATANOY 151
 DB 77 NNRKAVDQ--AGYVNFAYIEQTGNANDASISQSYAGNSAIIIGKSGNKANITQY 129

RESULT 14

AH0635
 nucleation component of curlin monomers [imported] - *Salmonella enterica* subsp. *enterica*
 C/Species: *Salmonella enterica* subsp. *enterica* serovar *Typhi*
 A/Note: this species has also been called *Salmonella typhi*
 C/Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
 C/Accession: AH0635

R/Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher
 et al.; S.; Mout, S.; O'Garra, P.
 Nature 413, 848-852, 2001
 A/Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skellon, J.; Stevens, K.
 A/Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serc
 A/Reference number: AB0502; MUID:21534947; PMID:11677608
 A/Accession: AH0635
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-151 <PAR>
 A/Cross-references: GB:AL513382; PIDN:CAD08267.1; PID:g16502314; GSPDB:GN00176
 C/Genetic8:
 A/Gene: STY1180

Query Match 13.7%; Score 107; DB 2; Length 151;
 Best Local Similarity 30.4%; Pred. No. 0.039;
 Matches 35; Conservative 17; Mismatches 47; Indels 16; Gaps 5;

QY 51 GSANALYD--QLVTRVTHEMAHAGYNGADVQ-GADNST-----IELTNGF 97
 DB 17 GATATNTYDLARSEYNFAVELSKSSFNQAIIIGQVGTDSARVROGSKLSVTSQEG 76
 QY 98 RNNATIDQWNAKNSDITVQYQGSANALVNOTASDSSVMVRYQYFGNNATANOY 151
 DB 77 NNRKAVDQ--AGYVNFAYIEQTGNANDASISQSYAGNSAIIIGKSGNKANITQY 129

RESULT 15

JC2143
 ice nucleation active protein - *Erwinia uredovora* (strain KUIN-3)
 C/Species: *Erwinia uredovora*
 C/Date: 30-Sep-1993 #sequence_revision 20-Aug-1994 #text_change 17-Mar-1999
 C/Accession: JC2143
 R/Michigami, Y.; Watabe, S.; Abe, K.; Ohta, H.; Arai, S.
 Biotechnol. Biochem. 58, 762-764, 1994
 A/Title: Cloning and sequencing of an ice nucleation active gene of *Erwinia uredovora*.
 A/Reference number: JC2143; MUID:94264407; PMID:7764866
 A/Accession: JC2143
 A/Molecule type: DNA
 A/Residues: 1-1034 <MIC>
 C/Comment: This protein consists of 52 repeats of closely related 16-amino acid motifs
 C/Superfamily: ice nucleation protein
 F/161-993/Region: R-domain

Query Match 13.6%; Score 106; DB 2; Length 1034;
 Best Local Similarity 29.9%; Pred. No. 0.39; 37; Indels 38; Gaps 9;
 Matches 40; Conservative 19; Mismatches 37; Indels 38; Gaps 9;

QY 27 GGGGNNGGNSGSPDSTLSIYQGSANALYDQVTRVVTHEMAHAGYNGADVQCAD 86
 DB 531 GYGSTAGANS-----LIAGYSTQTASVNSVLT-----AGVSTQTAREGSD 575
 QY 87 NSTIELTNGFRNATIDQWNAKNSDITVQYQ-----GNNALV-----NOTASDSSVM 136
 DB 576 -----LTA-GYGSTQTAOR-----NSDLTTG-YGSGTAGYSSLIAGYSGTQTAGYHSIL 624
 QY 137 VRQYFGNNATANO 150
 DB 625 T--AGYSGTQTAOR 636

Search completed: March 11, 2004, 18:42:11
 Job time : 10.3 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Comugen Ltd.

OM protein - protein search, using SW model

Run on: March 11, 2004, 16:14:48 ; Search time 6.3 Seconds
(without alignments)
1248.031 Million cell updates/sec

Title: US-09-543-407-26

Perfect score: 782
Sequence: 1 MKLLKYAAPAAIVSGSALA.....DSSVMYRGVGNMATAQY 151

Scoring table: BL0SUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	692	88.5	151	1	CSGA_SALTY
2	523	66.9	151	1	CSGA_SALTY
3	501.5	64.1	152	1	CSGA_SALTY
4	112	14.3	151	1	CSGA_SALTY
5	107.5	13.7	1322	1	ICEN_PANAN
6	107	13.7	151	1	CSGA_SALTY
7	107	13.7	151	1	CSGA_SALTY
8	106	13.6	1034	1	ICEN_PANAN
9	104.5	13.4	1258	1	ICEN_PANAN
10	102.5	13.1	1567	1	ICEN_PANAN
11	102.5	13.1	1567	1	ICEN_PANAN
12	97.5	12.5	1028	1	ICEN_PANAN
13	96.5	12.3	1196	1	ICEN_PANAN
14	95.5	12.2	1656	1	ICEN_PANAN
15	94	12.0	491	1	ICEN_PANAN
16	93	11.9	491	1	ICEN_PANAN
17	91.5	11.7	894	1	ICEN_PANAN
18	91	11.6	678	1	ICEN_PANAN
19	90.5	11.6	1200	1	ICEN_PANAN
20	89.5	11.4	1317	1	ICEN_PANAN
21	88.5	11.3	646	1	ICEN_PANAN
22	88.5	11.3	1148	1	ICEN_PANAN
23	88.5	11.3	1210	1	ICEN_PANAN
24	87.5	11.2	602	1	ICEN_PANAN
25	86.5	11.1	590	1	ICEN_PANAN
26	86	11.0	959	1	ICEN_PANAN
27	85	10.9	681	1	ICEN_PANAN
28	84.5	10.8	641	1	ICEN_PANAN
29	84.5	10.8	1300	1	ICEN_PANAN
30	84.5	10.8	1654	1	ICEN_PANAN
31	82.5	10.5	487	1	ICEN_PANAN
32	82.5	10.5	1327	1	ICEN_PANAN
33	82	10.5	1327	1	ICEN_PANAN

34	81.5	10.4	347	1	MSA2_PLA2
35	81.5	10.4	1778	1	N189_SCHPO
36	81	10.4	365	1	ROAI_DROME
37	81	10.4	443	1	SEU_ARATH
38	81	10.4	877	1	SPT_YEAST
39	80.5	10.3	1063	1	MLE_DROME
40	80	10.2	1293	1	SGG_DROME
41	79.5	10.2	1067	1	GRPI_ORYSA
42	79	10.1	165	1	RO21_XENLA
43	79	10.1	346	1	FRPA_NEIMB
44	79	10.1	1302	1	OMP_RICCT
45	79	10.1	1645	1	OMP_RICCT

ALIGNMENTS

RESULT 1	CSGA_SALTY	STANDARD	PRT	151 AA.
AC	P55225;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	10-OCT-1996 (Rel. 34, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Major curlin subunit precursor (Pimbrin SFF17).			
GN	CSGA OR AGFA OR STM1144 OR STY1181 OR T1776.			
OS	Salmonella typhimurium,			
OS	Salmonella typhi, and			
OC	Salmonella enteritidis.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OC	Enterobacteriaceae; Salmonella.			
NC	NCBI_Taxid=602, 601, 592;			
RM	[1]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=S.typhimurium; STRAIN=SR-11;			
RA	MEDLINE=98117058; PubMed=9457880;			
RA	Romling U., Bian Z., Hammar M., Sierralta W.D., Normark S.;			
RT	"Curli fibers are highly conserved between Salmonella typhimurium and			
RT	Escherichia coli with respect to operon structure and regulation.";			
RT	J. Bacteriol. 180:722-731(1998).			
RL	[2]			
RN	SEQUENCE FROM N.A.			
RP	SPECIES=S.typhimurium; STRAIN=LT2 / SSGC1412 / ATCC 700720;			
RC	MEDLINE=21534948; PubMed=11677609;			
RA	McClelland M., Sanderson K.R., Spieth J., Clifton S.W., Latreille P.,			
RA	Courney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,			
RA	Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,			
RA	Ryan B., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,			
RA	Waterston R., Wilson R.K.;			
RT	"Complete genome sequence of Salmonella enterica serovar Typhimurium			
RT	LT2.";			
RT	Nature 413:852-856(2001).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=S.typhi; STRAIN=CT18;			
RC	MEDLINE=21534947; PubMed=11677608;			
RA	Parthill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,			
RA	Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,			
RA	Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,			
RA	Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,			
RA	Feltwell T., Hamlin N., Haque A., Hien T.T., Holtroyd S., Jagels K.,			
RA	Krogh A., Larsen T.S., Leather S., Moule S., O'Gara P., Parry C.,			
RA	Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,			
RA	Whithead S., Barrett B.G.;			
RT	"Complete genome sequence of a multiple drug resistant Salmonella			
RT	enterica serovar Typhi CT18.";			
RT	Nature 413:848-852(2001).			
RL	[4]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=S.typhi; STRAIN=TY2 / ATCC 700931;			
RC	MEDLINE=22531367; PubMed=12644504;			
RX	Deng W., Liu S.-R., Plunkett G., III, Mayhew G.F., Rose D.J.,			
RA	Burkland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;			

"Comparative genomics of *Salmonella enterica* serovar Typhimurium strains Ty2 and CT18.";
 RL J. Bacteriol. 185:2330-2337(2003).
 RN [5]
 RP SEQUENCE FROM N.A.
 RX SPECIES=S.enteritidis; STRAIN=27655-3B;
 RM MEDLINE=96146512; PubMed=8550497;
 RA Collinson S.K., Cloutier S.C., Doran J.L., Baner P.A., Kay W.W.;
 RT "Salmonella enteritidis agfA/C operon encoding thin, aggregative
 fimbriae";
 RL J. Bacteriol. 178:662-667(1996).
 RN [6]
 RP SEQUENCE OF 21-151 FROM N.A.
 RX SPECIES=S.enteritidis; STRAIN=27655-3B;
 RM MEDLINE=94013373; PubMed=8104955;
 RA Doran J.L., Collinson S.K., Burian J., Santos G., Todd E.C.D.,
 RM Munro C.K., Kay C.M., Baner P.A., Peterkin P.I., Kay W.W.;
 RT "DNA-based diagnostic tests for *Salmonella* species targeting agfA,
 the structural gene for thin, aggregative fimbriae";
 RL J. Clin. Microbiol. 31:2263-2273(1993).
 RN [7]
 RP SEQUENCE OF 21-33.
 RX SPECIES=S.enteritidis; STRAIN=27655-3B;
 RM MEDLINE=91310586; PubMed=1677357;
 RA Collinson S.K., Emeody L., Mueller K.-W., Trust T.J., Kay W.W.;
 RT "Purification and characterization of thin, aggregative fimbriae from
Salmonella enteritidis";
 RL J. Bacteriol. 173:4773-4781(1991).
 CC -I- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 FIBRONECTIN.
 CC -I- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see http://www.isb-sib.ch/announce/
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AJ002301; CA05317.1; -
 DR EMBL; AE008749; AL20074.1; -
 DR EMBL; AF627269; CAD08268.1; -
 DR EMBL; AE016840; AAC693399.1; -
 DR EMBL; U43280; AAC43599.1; -
 DR PIR; JC6039; JC6039.
 DR StGene; SG10608; csgA.
 DR Fimbrin; Signal; Complete proteome.
 FT CHAIN 1 20
 FT SIGNAL 1 20
 FT CHAIN 21 151 MAJOR CURLIN SUBUNIT.
 FT CONFLICT 134 151 SVNVROYFGNNATANOY -> DSYTOVAS (IN
 REF. 6).
 SO SEQUENCE 151 AA; 15305 MM; BYDACOD16B21359 CRC64;
 QY Query Match 88.5%; Score 692; DB 1; Length 151;
 Best Local Similarity 90.7%; Pred. No. 3,2e-51;
 Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;
 OY 1 MKLKVAFAAIIVSGSALAGVVPQWGCGGNGHNSGGSDSTLSIYOGSANAAALYDQ 60
 Db 1 MKLKVAFAAIIVSGSALAGVVPQWGCGGNGHNSGGSDSTLSIYOGSANAAALALD 60
 OY LVTVVTHMAHAGYGADVGADGSTETLTONGFRNNAITTDOMAKNSDITTYOYXG 120
 Db 61 SDARKSETTTIGSIGAGADVGGGADNSTLETTONGFRRNATTIDOMAKNSDITTYOYXG 120
 OY 121 NNAAVLVQTASDDSVMYROYGFGNNATANYOY 151
 Db 121 NNAAVLVQTASDDSVMYROYGFGNNATANYOY 151

RESULT 2
 CSGA_ECOLI STANDARD; PERT; 151 AA.
 ID _CSGA_ECOLI
 AC P28307;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 26-FEB-2003 (Rel. 41, Last annotation update)
 DE Major curlin subunit precursor.
 GN CSGB OR B1042.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacterales;
 OC Enterobacteriaceae; Escherichia.
 NX NCBI_Taxid=562;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / W3110;
 RM MEDLINE=93211294; PubMed=8459772;
 RA Olsen A., Arngvist A.;
 RT "The RpoS sigma factor relieves H-NS-mediated transcriptional
 repression of csgA, the subunit gene of fibronectin-binding curli in
Escherichia coli";
 RL Mol. Microbiol. 7:523-536(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MC4100;
 RM MEDLINE=96414468; PubMed=8817489;
 RA Hammar M., Arngvist A., Bian Z., Olsen A., Normark S.;
 RT "Expression of two csg operons is required for production of
 fibronectin- and Congo red-binding curli polymers in *Escherichia coli*
 K-12";
 RL Mol. Microbiol. 18:661-670(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RM MEDLINE=97426617; PubMed=9278503;
 RA Blatner F.R., Plunkett G., III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Colado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor Y., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of *Escherichia coli* K-12";
 RL Science 277:1453-1474(1997).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RM MEDLINE=97061202; PubMed=8905232;
 RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizouchi K.,
 RA Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
 RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horiiuchi T.;
 RT "A 718-kb DNA sequence of the *Escherichia coli* K-12 genome
 corresponding to the 12.7-28.0 min region on the linkage map";
 RL DNA Res. 3:137-155(1996).
 RN [5]
 RP SEQUENCE OF 21-40.
 RC STRAIN=K12 / YMEJ;
 RM MEDLINE=93023873; PubMed=1357528;
 RA Arngvist A., Olsen A., Pfeiffer J., Russell D.G., Normark S.;
 RT "The Crl protein activates cryptic genes for curli formation and
 fibronectin binding in *Escherichia coli* HB101";
 RL Mol. Microbiol. 6:2443-2452(1992).
 RN [6]
 RP SEQUENCE OF 21-31.
 RX MEDLINE=91310586; PubMed=1677357;
 RA Collinson S.K., Emeody L., Trust T.J., Kay W.W.;
 RT "Purification and characterization of thin, aggregative fimbriae from
Salmonella enteritidis";
 RL J. Bacteriol. 173:4773-4781(1991).
 CC -I- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 FIBRONECTIN.

```

CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L04979; AAA23616.1; -
DR EMBL; X90754; CAA62282.1; -
DR EMBL; AE000205; AAC74126.1; -
DR EMBL; D90741; BAA35832.1; -
DR EMBL; D90742; BAA35840.1; -
DR PIR; S70788; S70788.
DR EcoGene; EG11489; csga.
KM Rimbria; Signal; Complete proteome.
FT SIGNAL 1 20
FT CHAIN 21 151 MAJOR CURLIN SUBUNIT.
FT CONFLICT 7 7 A -> B (IN REF. 1).
SQ SEQUENCE 151 AA; 15049 MW; C003470D208D395F CRC64;

Query Match 66.9%; Score 523; DB 1; Length 151;
Best Local Similarity 68.9%; Pred. No. 4.1e-37;
Matches 104; Conservative 18; Mismatches 29; Indels 0; Gaps 0;

QY 1 MKLLKVAFAFAIVSGSALAGVWPWGSGNHNHGGSSGPDSTLTSTIYOGSANAALYDQ 60
DB 1 MKLLKVAFAFAIVSGSALAGVWPWGSGNHNHGGSSGPDSTLTSTIYOGSANAALQ 60
QY 61 LVTRVVTHEMAHAGYNGADVGGADNSTIELTONGFRNNATIDOMAKNSDITVGGYG 120
DB 61 TDARNSDLTITGCGGAGADVGGSDSDSIDLTGRCGNSATLIDQNGKSEMTVQFG 120
QY 121 NNALVNOTASDSSVMRQVFGNNATANOY 151
DB 121 GNGAAYDQTASNSVTNVTVQVFGNNATANOY 151

RESULT 3
CSGB_ECO57 STANDARD; PRT; 152 AA.
ID CSGA_ECO57 STANDARD; PRT; 152 AA.
AC 093U24;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Major curlin subunit precursor.
GN CSGA OR Z1676 OR ECS1420.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / ATCC 43895;
RX MEDLINE=21218556; PubMed=11319125;
RA Uhlrich G.A., Keen J.E., Elder R.O.;
RT "Mutations in the csq promoter associated with variations in curli
RL expressions in certain strains of Escherichia coli O157:H7.";
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G., Burland V., Mau B., Glasner J.D.,
RA Rose D.U., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner P.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).

```

```

RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 050952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,
RA Kihara S., Shiba T., Hatcori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RL O157:H7 and genomic comparison with a laboratory strain K-12.";
RN DNA Res. 8:11-22(2001)
CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN.
CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF275733; AAK53212.1; -
DR EMBL; AE005315; AAG55788.1; -
DR EMBL; AP002554; BAB34843.1; -
DR PIR; D90806; D90806.
DR PIR; H85665; H85665.
KM Rimbria; Signal; Complete proteome.
FT SIGNAL 1 20 BY SIMILARITY.
FT CHAIN 21 152 MAJOR CURLIN SUBUNIT.
SQ SEQUENCE 152 AA; 15099 MW; EE2D2D94DDE91243 CRC64;

Query Match 64.1%; Score 501.5; DB 1; Length 152;
Best Local Similarity 67.1%; Pred. No. 2.6e-35;
Matches 102; Conservative 19; Mismatches 30; Indels 1; Gaps 1;

QY 1 MKLLKVAFAFAIVSGSALAGVWPWGSGNHNHGGSSGPDSTLTSTIYOGSANAALYD 59
DB 1 MKLLKVAFAFAIVSGSALAGVWPWGSGNHNHGGSSGPDSTLTSTIYOGSANAALQ 60
QY 60 QLVTRVVTHEMAHAGYNGADVGGADNSTIELTONGFRNNATIDOMAKNSDITVGGYG 119
DB 60 QDARNSDLTITGCGGAGADVGGSDSDSIDLTGRCGNSATLIDQNGKSEMTVQFG 120
QY 120 GNNALVNOTASDSSVMRQVFGNNATANOY 151
DB 121 GNGAAYDQTASNSVTNVTVQVFGNNATANOY 152

RESULT 4
CSGB_ECOLI STANDARD; PRT; 151 AA.
ID CSGA_ECOLI STANDARD; PRT; 151 AA.
AC P39828;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Minor curlin subunit precursor.
GN CSGA OR B1041 OR Z1675 OR ECS1419.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562; 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MC4100;
RX MEDLINE=96414468; PubMed=8817489;
RA Hammar M., Arqvist A., Bian Z., Olsen A., Normark S.;
RT "Expression of two csq operons is required for production of
RL fibronectin- and congo red-binding curli polymers in Escherichia coli

```

RT K-12." ;
ML Microbiol. 18:661-670(1995).
[2]
SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of *Escherichia coli* K-12." ;
RL Science 277:1453-1474(1997).
[3]
SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061302; PubMed=8905232;
RA Oshima T., Alta H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kasahimoto K.,
RA Kimura S., Kitagawa M., Makino K., Maeda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiiuchi T.;
RT A 718-kb DNA sequence of the *Escherichia coli* K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map." ;
RL DNA Res. 3:137-155(1996).
[4]
SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Forstai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Pocumans K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7." ;
RL Nature 409:529-533(2001).
[5]
SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Saekawa C., Ogasawara N., Yasunaga T.,
RA Kuhnara S., Shiba T., Hattori M., Shiragawa H.;
RT "Complete genome sequence of enterohaemorrhagic *Escherichia coli*
RT O157:H7 and genomic comparison with a laboratory strain K-12." ;
RL DNA Res. 8:11-22(2001).
[6]
SEQUENCE OF 1-21 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=95157246; PubMed=7854117;
RA Arrgvast A., Olsen A., Normark S.;
RT "Sigma S-dependent growth-phase induction of the *csGA* promoter in
RT *Escherichia coli* can be achieved in vivo by sigma 70 in the absence
RT of the nucleoid-associated protein H-NS." ;
ML Microbiol. 13:1021-1032(1994).
-1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC FIBROECRINS BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC CURLIN MONOMERS.
CC
CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL, X90754, CAA62281.1, -.

```

DR EMBL: AE000205; AAC74125.1; -
DR EMBL: D90741; BAA35831.1; -
DR EMBL: AE005315; AAG55787.1; -
DR EMBL: AP002554; BAB34842.1; -
DR PIR: G90806; C90806.
DR PIR: G85665; G85665.
DR PIR: S70787; S70787.
DR EcoGene: EG1621; csgB.
KW Fimbria; Signal; Complete proteome.
FT SIGNAL
FT CHAIN 1 21 POTENTIAL.
SQ SEQUENCE 151 AA; 15882 MW; B18D26B964014B8 CRC64;

Query Match 14.3%; Score 112; DB 1; Length 151;
Best Local Similarity 29.8%; Pred. No. 0.0083;
Matches 34; Conservative 17; Mismatches 49; Indels 14; Gaps 3

Oy 51 GSANALYDOLVT--RYVTHMAHAGYGNADVQAGADNSTIELTONGFR----- 98
Db 17 GIAAAYDYDLANSYFNVELSKSFQAIIIOAGTNNNSAQLRGGSKLAAVAQDGS 76

Oy 99 -NNATIDQWMAKNSDITVGOYGANNALVNTQASDSSVMVQVFGNNATANQY 151
Db 77 SNRAKIITQGTQDYNL-AVTDQAGSANDSISGATGNTAMTIQKSGNKANITQY 129

RESULT 5
ICEA PANAN STANDARD; PRT; 1322 AA.
ID ICEA PANAN
AC P20463;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ice nucleation protein inaA.
GN inaA.
OS Pantoea ananas (Erwinia uredovora).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pantoea.
OX NCBI_TaxID=553;
RX MEDLINE=90092494; PubMed=2599095;
RA Abe K, Watabe S, Emori Y, Watanabe M, Arai S.;
RT "An ice nucleation active gene of Erwinia ananas. Sequence similarity
RT to those of Pseudomonas species and regions required for ice
RT nucleation activity."
RL FEBS Lett. 258:297-300(1989).
CC -1- FUNCTION: Ice nucleation proteins enable bacteria to nucleate
CC -1- crystallization in supercooled water.
CC -1- SUBCELLULAR LOCATION: Outer membrane (By similarity).
CC -1- DOMAIN: CONTAINS MANY IMPERFECT REPEATS OF THE CONSENSUS
CC OCTAPEPTIDE A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A
CC REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.
CC -1- SIMILARITY: Belongs to the bacterial ice nucleation protein
CC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.1ab-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: X17316; CAA35194.1; -.
DR PIR: S07053; S07053.
DR HSSP: P06620; IINA.
DR InterPro: IPR000258; Ice_nucleatn.
DR Pfam: PF00818; Ice_nucleatn; 69.
DR PRINTS: PR00327; ICENUCLEATN.
DR PROSITE: PS00314; ICE_NUCLEATION; 49.
DR Ice_nucleation/Repeat; Outer membrane.
KW DOMAIN 162 1281 OCTAPEPTIDE PERIODICITY.

```

50 SEQUENCE 1322 AA; 131094 MW; 8990E24AA837039 CRC64;
 Query Match 13.7%; Score 107.5; DB 1; Length 1322;
 Best Local Similarity 29.9%; Pred. No. 0.22;
 Matches 38; Conservative 19; Mismatches 21; Indels 49; Gaps 8;
 QY 34 GCGSSGGDTSLTGYGSAANAALYDQVTRVTHMAHAGYGCAGDVGADNSTIELT 93
 DB 933 GSTAGPDSLSL-IAGYSTQTAGYNSILT-----AGYGS-----T 967
 QY 94 QNGFRNNTIDQWNAKNSDITVGOY-----GNNALY-----NOTASDSSVMRQVGF 143
 DB 968 QTC-----QENSDLTTG-YGSTTAGYESSLIAGSTQTAGFKSTLM--AGYV 1013
 QY 144 NNATAMQ 150
 DB 1014 SSQTARE 1020
 RESULT 6
 CSGB_SALTY STANDARD; PRT; 151 AA.
 AC 0827M3:
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Minor curlin subunit precursor.
 GN CSGB OR STY1180 OR T1777.
 OS Salmonella typhi.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 OX NCBI_TaxID=601;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CT18:
 RX MEDLINE=21534947; PubMed=11677608;
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
 RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaiha M.,
 RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
 RA Felwell T., Hamlin N., Hogue A., Hien T.T., Holtroff S., Jajula K.,
 RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Patry C.,
 RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
 RA Whitehead S., Barrett B.G.;
 RT "Complete genome sequence of a multiple drug resistant Salmonella
 RT enterica serovar Typhi CT18.";
 RL Nature 413:848-852(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Ty2 / ATCC 700931;
 RX MEDLINE=22531367; PubMed=12644504;
 RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
 RA Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;
 RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
 RT and CT18.";
 RL J. Bacteriol. 185:2330-2337(2003).
 CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
 CC CURLIN MONOMERS.
 CC -1- SIMILARITY: BELONGS TO THE CSGB/CSGB FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AL627269; CA008267.1; -
 DR EMBL; AE016840; AA069400.1; -

KW Fimbria; Signal; Complete proteome.
 FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 22 151 MINOR CURLIN SUBUNIT.
 50 SEQUENCE 151 AA; 16254 MW; 161C5432E573495 CRC64;
 Query Match 13.7%; Score 107; DB 1; Length 151;
 Best Local Similarity 30.4%; Pred. No. 0.022;
 Matches 35; Conservative 17; Mismatches 47; Indels 16; Gaps 5;
 QY 51 GSANAALYD--QLVTRVTHMAHAGYGCAGDVGQ--GADNST-----IELTNGF 97
 DB 17 GIATRYVDLARSTYNAVNBLSKSPNOAIIQVGTIDNSARVRQSGSKLSVISDGE 76
 QY 98 RNNATIDQWNAKNSDITVGOYGNNALYNOTASDSSVMRQVGFNNATAMQY 151
 DB 77 NNRAKVDQ--AGVNFAYIETGTGNANDASISQSAVGSAAIIQKSGNKANITQY 129
 RESULT 7
 CSGB_SALTY STANDARD; PRT; 151 AA.
 AC P55226:
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Minor curlin subunit precursor (Fimbrin SFP17, minor subunit).
 GN CSGB OR AGFB OR STM1143.
 OS Salmonella typhimurium, and
 OS Salmonella enteritidis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 OX NCBI_TaxID=602, 592;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S.typhimurium; STRAIN=SR-11;
 RX MEDLINE=98117058; PubMed=9457880;
 RA Romling U., Bian Z., Hammar M., Sierraalta W.D., Normark S.;
 RT "Curli fibers are highly conserved between Salmonella typhimurium and
 RT Escherichia coli with respect to operon structure and regulation.";
 RL J. Bacteriol. 180:722-731(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S.typhimurium; STRAIN=L72 / SGSC1412 / ATCC 700720;
 RX MEDLINE=21534948; PubMed=11677609;
 RA McEllelland M., Sanderson K.B., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Portolillo S., Ali J., Dante M., Du J., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
 RA Ryan B., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
 RT L72.";
 RL Nature 413:852-856(2001).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S.enteritidis; STRAIN=27655-3B;
 RX MEDLINE=96146512; PubMed=8550497;
 RA Collinson S.K., Cloutier S.C., Doran J.L., Baner P.A., Kay W.W.;
 RT "Salmonella enteritidis agfBAC operon encoding thin, aggregative
 RT fimbriae.";
 RL J. Bacteriol. 178:662-667(1996).
 CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
 CC CURLIN MONOMERS.
 CC -1- SIMILARITY: BELONGS TO THE CSGB/CSGB FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

```

CC or send an email to license@ib-sib.ch).
CC -----
CC EMBL: AJ002301; CA05316.1; -.
CC DR EMBL: AE008749; AL20073.1; -.
CC DR EMBL: U43280; JAC43598.1; -.
CC DR PIR: JG6040; JG6040.
CC DR Stygen; SG10609; csgB.
CC KW Pimbrja; Signal; Complete proteome.
CC FT SIGNAL 1 21 POTENTIAL.
CC FT CHAIN 151 MINOR CURLIN SUBUNIT.
CC SQ SEQUENCE 151 AA; 16182 MW; C0FC5430E6DD361D CRC64;

Query Match
Best Local Similarity 30.4%; Pred. No. 0.022; DB 1; Length 151;
Matches 35; Conservative 17; Mismatches 47; Indels 16; Gaps 5;

QY 51 GSANALYD-OLVTRVTHENAHHAGYGADVGQ-GADNST-----IELTNGF 97
DB 17 GIATATNTVDLASENFANFANELSKSFNQALIGVGTDNSARVQESKLSVTSQSG 76
QY 98 RNNAITDQNAKNSDIT-VGYGANNALVNOTASDSSVMRQVFGNNTATNOY 151
DB 77 NNRKAVDQ--AGNYNFAYIEGTGNANDASISQAVGNSMAITQKSGNKANITQY 129

RESULT 8
ID ICEN_PANAN STANDARD; PRT; 1034 AA.
AC Q47879;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ice nucleation protein inau.
GN INAU.
OS Pantoea ananas (Erwinia uredovora).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pantoea.
OX NCBI_TaxID=553;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KUN-3;
RX MEDLINE=94264407; PubMed=7764866;
RA Michigami Y., Watabe S., Abe K., Obata H., Arai S.;
RT "Cloning and sequencing of an ice nucleation active gene of Erwinia
RT uredovora.";
RL Biosci. Biotechnol. Biochem. 58:762-764(1994).
CC -1- FUNCTION: Ice nucleation proteins enable bacteria to nucleate
CC crystallization in supercooled water.
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- DOMAIN: CONTAINS IMPERFECT REPEATS OF A CONSENSUS OCTAPEPTIDE
CC A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A REGIONAL 48-RESIDUE
CC PERIODICITY IS SUPERIMPOSED.
CC -1- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE
CC NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.
CC -1- SIMILARITY: Belongs to the bacterial ice nucleation protein
CC family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ibt-sib.ch/announce/
CC or send an email to license@ib-sib.ch).
CC -----
CC EMBL: D14992; BAA03636.1; -.
CC DR PIR: JC2143; JC2143.
CC DR HSSP: P06620; 11NA.
CC DR InterPro: IPR000258; Ice_nucleatn.
CC DR Pfam: PF00818; Ice_nucleatn; 51.
CC DR PRINTS: PR00327; ICENUCLEATN.
CC DR PROSITE: PS00314; ICE_NUCLEATION; 34.
CC KW Ice nucleation; Repeat; Outer membrane.

```

```

FT DOMAIN 162 993 OCTAPEPTIDE PERIODICITY.
SQ SEQUENCE 1034 AA; 103378 MW; FA225232D333EADD CRC64;

Query Match
Best Local Similarity 29.9%; Pred. No. 0.22;
Matches 40; Conservative 19; Mismatches 37; Indels 38; Gaps 9;

QY 27 GGGNNHNGGSGSDSTLSIYOGSANAALYDQVTRVTHENAHHAGYGADVGQAD 86
DB 531 GYGSTSTAGANS-----LIAGYSTQTASINSVLT-----AGYSTQTAREGSD 575
QY 87 NSTIELTNGFRNNAITDQNAKNSDITVQYG-----GNNALV-----NQIASDSSVM 136
DB 576 -----LTA-GYGSTQTAGE-----NSDLTTG-YGSTSTAGYDSSLIAGYSTQTGYHSIL 624
QY 137 VRQVFGNNTATNQ 150
DB 625 T--AGYSTQTAGE 636

RESULT 9
ID ICEN_ERWHE STANDARD; PRT; 1258 AA.
AC P16239;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ice nucleation protein.
GN ICE.
OS Erwinia herbicola.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pantoea.
OX NCBI_TaxID=549;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M1;
RX MEDLINE=90152370; PubMed=2515997;
RA Warren G.J., Corotto L.V.;
RT "The consensus sequence of ice nucleation proteins from Erwinia
RT herbicola, Pseudomonas fluorescens and Pseudomonas syringae.";
RL Gene 85:239-242(1989).
CC -1- FUNCTION: Ice nucleation proteins enable bacteria to nucleate
CC crystallization in supercooled water.
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- DOMAIN: CONTAINS 126 IMPERFECT REPEATS OF A CONSENSUS OCTAPEPTIDE
CC A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A REGIONAL 48-RESIDUE
CC PERIODICITY IS SUPERIMPOSED.
CC -1- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE
CC NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.
CC -1- SIMILARITY: Belongs to the bacterial ice nucleation protein
CC family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ibt-sib.ch/announce/
CC or send an email to license@ib-sib.ch).
CC -----
CC EMBL: M26382; AAA24823.1; -.
CC DR PIR: JQ0188; JQ0188.
CC DR HSSP: P06620; 11NA.
CC DR InterPro: IPR000258; Ice_nucleatn.
CC DR Pfam: PF00818; Ice_nucleatn; 65.
CC DR PRINTS: PR00327; ICENUCLEATN.
CC DR PROSITE: PS00314; ICE_NUCLEATION; 45.
CC KW Ice nucleation; Repeat; Outer membrane.
CC DOMAIN 162 1217 OCTAPEPTIDE PERIODICITY.
CC SQ SEQUENCE 1258 AA; 125084 MW; 590E8A130077BD4 CRC64;

Query Match
Best Local Similarity 29.5%; Pred. No. 0.36;

```


Matches 44; Conservative 23; Mismatches 37; Indels 45; Gaps 10;

QY 34 GCGSSGSPSTLSIYOGSANAALYDQVTVVTHEMAHAGYN-----GA 79
 DB 773 GSTGTAGADSSIL-IAGYGTGTAGYHSILT-----AGYGTGTAGERRSDLTGYS 822
 QY 80 DVGGAGDNSTIE---LTQN-GFRNNATI-----DOMNAKSDIVGVGY-----GNNAAIV 126
 DB 823 TSTAGADSSLIAGYGTGTAGYHSILTLAGYGTGTAGYHSILTTG-YGSTGTAGYDSSL 881
 QY 127 -----NOTASDSSVMVROVGFNNATAMQ 150
 DB 882 AGYGTGTAGYHSILT--AGYGTGTAGAE 908

RESULT 10
 ICEN_XANCT STANDARD; PRT; 1567 AA.
 AC P18127;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Ice nucleation protein.
 GN INAX.
 OS Xanthomonas campestris (pv. translucens).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 OC Xanthomonadaceae; Xanthomonas.
 OX NCBI_TaxID=343;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=X565;
 RX MEDLINE=91080859; PubMed=2259339;
 RA Zhao J., Orser C.S.;
 RT "Conserved repetition in the ice nucleation gene *inx* from
 RT *Xanthomonas campestris* pv. *translucens*.";
 RL Mol. Gen. Genet. 223:163-166(1990).
 CC -1- FUNCTION: Ice nucleation proteins enable bacteria to nucleate
 CC crystallization in supercooled water.
 CC -1- SUBCELLULAR LOCATION: Outer membrane (By similarity).
 CC -1- DOMAIN: CONTAINS 153 IMPERFECT REPEATS OF THE CONSENSUS
 CC OCTAPEPTIDE A-G-Y-G-S-T-L-T. FURTHER ON A 16-RESIDUE AND A
 CC REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.
 CC -1- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE
 CC NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.
 CC -1- SIMILARITY: Belongs to the bacterial ice nucleation protein
 CC family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL; X52970; CAA37140.1; -.
 DR HSSP; P06620; 11NA.
 DR InterPro; IPR000258; Ice_nucleatn.
 DR Pfam; PF00818; Ice_nucleatn; 81.
 DR PRINTS; PR00327; ICENUCLEATIN.
 DR PROSITE; PS00314; ICE_NUCLEATION; 57.
 KW Ice nucleation; Repeat; Outer membrane.
 SQ SEQUENCE 1567 AA; 152548 MW; CBB451D959ECAD63 CRC64;

Query Match 13.1%; Score 102.5; DB 1; Length 1567;
 Best Local Similarity 30.2%; Pred. No. 0.68; Indels 45; Gaps 11;
 Matches 45; Conservative 23; Mismatches 36; Indels 45; Gaps 11;

QY 34 GCGSSGSPSTLSIYOGSANAALYDQVTVVTHEMAHAGYN-----GADV----- 81
 DB 1055 GSTGTAGADSSIL-IAGYGTGTAGYDSSSLT-----AGYGTGTAGGSDITAGYGS 1104
 QY 82 --GCGADNSTIE---LTQN-GFRNNATI-----DOMNAKSDITGVGY-----GNNAAIV 126

DB 1105 TGTAGADSSLIAGYGTGTAGYDSSNLITAGYGTGTAGERRSDLTAG-YGSTGTAGHSSSL 1163
 QY 127 -----NOTASDSSVMVROVGFNNATAMQ 150
 DB 1164 AGYGTGTAGYHSILT--TGYSTGTAGAE 1190

RESULT 11
 OMPB_RICCN STANDARD; PRT; 1655 AA.
 ID OMPB_RICCN
 AC O9KKK3; O9KK98; O9XC45;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Outer membrane protein B precursor (168 kDa surface-layer protein)
 DE (Surface protein antigen) (Cell surface antigen 5) (Scas) (rOmpB)
 DE (rOmp B) [contains: 120 kDa surface-exposed protein (Surface protein
 DE antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide].
 GN OMPB OR RC1085.
 OS Rickettsia conorii.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 OX NCBI_TaxID=781;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Malish 7;
 RX MEDLINE=21442074; PubMed=11557893;
 RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-B., Barbe V.,
 RA Samson D., Roux V., Cosset P., Weissenbach J., Claverie J.-M.,
 RA Raoult D.;
 RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
 RL Science 293:2093-2098(2001).
 RN [2]
 RP SEQUENCE OF 33-1649 FROM N.A.
 RC STRAIN=Indian tick typhus, and Malish 7;
 RX MEDLINE=20393643; PubMed=10939649;
 RA Roux V., Raoult D.;
 RT "Phylogenetic analysis of members of the genus Rickettsia using the
 RT gene coding the outer-membrane protein rOmpB (ompB).";
 RL Int. J. Syst. Evol. Microbiol. 50:1449-1455(2000).
 RN [3]
 RP SEQUENCE OF 353-1655 FROM N.A.
 RC STRAIN=Malish 7;
 RA Stenos J., Walker D.;
 RT "The rickettsial outer membrane protein A and B genes of Rickettsia
 RT australis, the most divergent rickettsia of the spotted fever group.";
 RL Submitted (MAY-1999) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
 CC VIRULANCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (By
 CC similarity).
 CC -1- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR
 CC (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-
 CC layer with hexagonal symmetry (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL; AE008659; AAL03623.1; -.
 DR EMBL; AF123721; AAF34124.1; -.
 DR EMBL; AF123726; AAF34129.1; -.
 DR EMBL; AF149110; AAD39533.1; -.
 DR PIR; E97835; E97835.
 DR InterPro; IPR006315; Autotransport.
 DR InterPro; IPR005546; Autotransporter.

DR Pfam; PF03797; Autotransporter; 1.
 DR TIGRfam; TIGR01414; autotrans. bart; 2.
 KW Antigen; S-layer; Cell wall; Complete proteome.
 FT CHAIN 1 1334 120 kDa SURFACE-EXPOSED PROTEIN.
 FT 1335 1655
 FT VARIANT 61 61 32 kDa BETA PEPTIDE.
 FT VARIANT 75 75 P -> A (IN STRAIN INDIAN TICK TYPHUS).
 FT VARIANT 78 78 G -> S (IN STRAIN INDIAN TICK TYPHUS).
 FT VARIANT 251 251 G -> N (IN STRAIN INDIAN TICK TYPHUS).
 FT VARIANT 413 413 N -> D (IN STRAIN INDIAN TICK TYPHUS).
 FT VARIANT 959 959 I -> V (IN STRAIN INDIAN TICK TYPHUS).
 FT VARIANT 968 968 A -> T (IN STRAIN INDIAN TICK TYPHUS).
 FT VARIANT 1139 1139 R -> L (IN STRAIN INDIAN TICK TYPHUS).
 FT CONFLICT 333 354 KO -> GH (IN REF. 3).
 FT CONFLICT 776 776 F -> S (IN REF. 3).
 FT CONFLICT 1159 1159 E -> D (IN REF. 3).
 FT CONFLICT 1177 1177 G -> S (IN REF. 3).
 FT CONFLICT 1492 1492 H -> R (IN REF. 3).
 SQ SEQUENCE 1655 AA; 168342 MW; E49E19377D5FC37 CRC64;

Query Match 13.1%; Score 102.5; DB 1; Length 1655;
 Best Local Similarity 26.8%; Pred. No. 0.73;
 Matches 38; Conservative 18; Mismatches 55; Indels 31; Gaps 6;

QY 28 GGGNNG--GNSGSDSTLSIYOGSANAALYDQVTR-----VTHMAHAGYNGA 79
 DB 1211 GGARFNGTLGGFNFVVTGSNRFVNYGLIRANQDVITRTNNAENVVTINDIANGSFGGAP 1270
 QY 80 DVGGAGADNSTELTNGFRNNAITDQWAKNSDITVGYGNNALVNGTSDSSWVR- 138
 DB 1271 GVGQNV-TTFVATVTAAYNNLL-----AKNS-----ANSANFGAIVTDSAITN 1317
 QY 139 -----QVFGNNATANY 151
 DB 1318 AQLDVAKDIOAQLGNRLGALRY 1339

RESULT 12
 ID OVO DROME STANDARD; PRT; 1028 AA.
 AC P51521; Q9XZU4;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE OVO protein (Shaven baby protein).
 GN OVO OR SVB.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Ovary;
 RX MEDLINE=95021209; PubMed=7935398;
 RA Garfinkel M.D., Wang J., Liang Y., Mahowald A.P.;
 RT "Multiple products from the Shavenbaby-ovo gene region of Drosophila
 melanogaster: relationship to genetic complexity.";
 RL Mol. Cell. Biol. 14:6809-6818(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Oregon-R;
 RX MEDLINE=9129102; PubMed=1712294;
 RA Mevel-Ninio M.T.M., Terracol R., Kafatos F.C.;
 RT "The ovo gene of Drosophila encodes a zinc finger protein required
 for female germ line development.";
 RL EMBO J. 10:2259-2266(1991).
 CC - FUNCTION: REQUIRED FOR SURVIVAL AND DIFFERENTIATION OF FEMALE GERM
 CC LINE CELLS. PLAYS A ROLE IN GERM LINE SEX DETERMINATION.
 CC - SUBCELLULAR LOCATION: Nuclear (Potential).
 CC - DEVELOPMENTAL STAGE: FIRST APPEARS IN THE GERMINAL AND
 CC ACCUMULATES IN NURSE CELLS DURING OOGENESIS. STORED IN THE EGG,
 CC BUT IS RAPIDLY LOST IN THE EMBRYOS EXCEPT FOR ITS CONTINUED

CC PRESENCE IN THE GERM LINE PRECURSOR POLE CELLS.
 CC - SIMILARITY: Contains 4 C2H2-type zinc fingers.
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb.ch).
 CC -----
 DR EMBL; U11383; AAB60216.1; -
 DR EMBL; X59772; CAB36921.1; ALT_SEQ.
 DR PIR; A56038; A56038.
 DR HSSP; P07248; 2ADR.
 DR TRANSFAC; T00669; -
 DR FLYBase; FBgn0003028; ovo.
 DR InterPro; IPR007087; ZnF_C2H2.
 DR Pfam; PF00096; zf-C2H2; 3.
 DR SMART; SM00355; ZnF_C2H2_4.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
 DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 3.
 KW Zinc-finger; Metal-binding; DNA-binding; Repeat; Nuclear protein;
 KW Transcription regulation.
 FT DOMAIN 62 66 POLY-ALA.
 FT DOMAIN 72 77 POLY-GLY.
 FT DOMAIN 80 85 POLY-GLY.
 FT DOMAIN 98 108 POLY-GLY.
 FT DOMAIN 144 152 POLY-HIS.
 FT DOMAIN 153 159 POLY-ASN.
 FT DOMAIN 336 339 POLY-GLN.
 FT DOMAIN 347 353 POLY-GLN.
 FT DOMAIN 357 361 POLY-GLN.
 FT DOMAIN 410 414 POLY-GLN.
 FT DOMAIN 418 422 POLY-GLN.
 FT DOMAIN 426 432 POLY-GLN.
 FT DOMAIN 445 453 POLY-GLN.
 FT DOMAIN 456 459 POLY-GLN.
 FT DOMAIN 466 474 POLY-GLN.
 FT DOMAIN 497 517 POLY-ALA.
 FT DOMAIN 524 529 POLY-SER.
 FT DOMAIN 549 558 POLY-ALA.
 FT DOMAIN 639 651 POLY-ALA.
 FT DOMAIN 717 725 POLY-ALA.
 FT DOMAIN 797 802 POLY-GLN.
 FT DOMAIN 820 823 POLY-GLN.
 FT DOMAIN 826 832 POLY-GLN.
 FT DOMAIN 874 886 C2H2-TYPE 1.
 FT ZN_FING 892 924 C2H2-TYPE 2.
 FT ZN_FING 930 953 C2H2-TYPE 3.
 FT ZN_FING 969 992 C2H2-TYPE 4.
 FT CONFLICT 647 647 A -> R (IN REF. 2).
 SQ SEQUENCE 1028 AA; 110620 MW; D7068B2BC0F677 CRC64;

Query Match 12.5%; Score 97.5; DB 1; Length 1028;
 Best Local Similarity 28.7%; Pred. No. 1.1;
 Matches 43; Conservative 12; Mismatches 60; Indels 35; Gaps 7;

QY 3 ILKVAFAPIAVISGALAGVWPWGCGGNHGGNSGSDSTLSIYOGSANAALYDQV 62
 DB 59 LONAAAPAIYISAGG-----GGGCTGGGGGASGGGPPANSGGGGGG----- 104
 QY 63 TRVTHMAHAGYNGADVGGAGADNSTELTNGFRNNAITDQWAKNSDI-----TVGQY 118
 DB 105 -----GGNGYINCQGVG-GPNN-----LDGNNLNPASVSNYNESNKRHHHHHHH 152
 QY 119 GGNNAALVNGTSDSSWVRVGFGNNATA 148
 DB 153 NNNN-----NNNGGQTSMGHPF-YGNDPSA 177

RESULT 13
 ICEV_PSEX

```

ID ICEV PSBSX STANDARD; PRT; 1196 AA.
AC 033479;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ice nucleation protein.
GN INAV.
OS Pseudomonas syringae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=317;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=INAV;
RX MEDLINE=97462815; PubMed=9323042;
RA Schmid D., Pridmore D., Capitani G., Battistutta R., Neeser J.-R.,
RA Jann A.;
RT "Molecular organization of the ice nucleation protein Inav from
RT Pseudomonas syringae.";
RL FEBS Lett. 414:590-594(1997).
CC -1- FUNCTION: Ice nucleation proteins enable bacteria to nucleate
CC crystallization in supercooled water.
CC -1- SUBUNIT: MEMBRANE ENVIRONMENT OR AGGREGATION SEEMS TO BE REQUIRED
CC FOR ICE NUCLEATION ACTIVITY.
CC -1- SUBCELLULAR LOCATION: Outer membrane (By similarity).
CC -1- DOMAIN: CONTAINS MANY IMPERFECT REPEATS OF THE CONSENSUS
CC OCTAPETIDE A-G-Y-G-S-T-L-T; FURTHER ON A 16-RESIDUE AND A
CC REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.
CC -1- SIMILARITY: Belongs to the bacterial ice nucleation protein
CC family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ001086; CA04521.1; -.
DR HSSP; P06620; 11NA.
DR InterPro; IPR000258; Ice_nucleatn.
DR Pfam; PF00818; Ice_nucleation; 61.
DR PRINTS; PR00327; ICNUCLEATN.
DR PROSITE; PS00314; ICE_NUCLEATION; 42.
KW Ice nucleation; Repeat; Outer membrane.
FT DOMAIN 172 1147 OCTAPETIDE PERIODICITY.
SQ SEQUENCE 1196 AA; 117991 MM; C9E9974CB1731B68 CRC64;

Query Match 12.3%; Score 96.5; DB 1; Length 1196;
Best Local Similarity 26.7%; Pred. No. 1.6;
Matches 43; Conservative 24; Mismatches 39; Indels 55; Gaps 11;

QY 16 GSALAG-----VVPQMG--GGNR-----GGNSGPDSTLSIYQGSANALYDOL 61
DB 175 GTLSGDNNRSLIAGYSGNETAGNSHDLIAGYSGTCTAGSDSL-VAGYSGTCTAGDGA 233
QY 62 VTRVVTHEMAHAGYGN-----GADVGGAGDNSTELTQNGFRNNATIDOWN 107
DB 234 LT-----AGYSGTCTARBSGSLTAGYSGTGTAGSDSLA-----GYGSIQT---S 276

QY 108 AKNSDITVGOYGGNNALVNTASDSSVWVROVFGNNATA 148
DB 277 GEDSLSLTAG-YGS-----TGTAGBSGSLT-AGYSGTGTGA 308

RESULT 14
OMP6 RICJA STANDARD; PRT; 1656 AA.
AC 006653;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

```

```

DE Outer membrane protein B precursor (168 kDa surface-layer protein)
DE (Surface protein antigen) (Cell surface antigen 5) (Scas) (rompB)
DE (romp B) (contains: 120 kDa surface-exposed protein (Surface protein
DE antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide).
GN OMP6.
OS Rickettsia japonica.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=35790;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YH;
RA Uchiyama T.;
RT "Sequencing of the gene encoding the protein romp B of Rickettsia
RT japonica.";
RL Submitted (May 1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THE 120 kDa SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (By
CC similarity).
CC -1- FUNCTION: THE 32 kDa BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR
CC (By similarity).
CC -1- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-
CC layer with hexagonal symmetry.
CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAE OMP6/OMP6 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB003681; BAA20138.1; -.
DR InterPro; IPR006315; Autotransport.
DR InterPro; IPR005546; Autotransporter.
DR Pfam; PF03797; Autotransporter; 1.
DR TIGRFAMs; TIGR01414; autotrans_bar1; 2.
KW Antigen; S-layer; Cell wall.
FT CHAIN 1 1338 120 kDa SURFACE-EXPOSED PROTEIN.
FT CHAIN 1339 1656 32 kDa BETA PEPTIDE.
FT FT 528 533 POLY-GLY.
SQ SEQUENCE 1656 AA; 168097 MM; 3132869C9DD5999F CRC64;

Query Match 12.2%; Score 95.5; DB 1; Length 1656;
Best Local Similarity 25.0%; Pred. No. 2.8;
Matches 36; Conservative 21; Mismatches 52; Indels 35; Gaps 7;

QY 28 GGNHNGGNSGPDSTLS---IYQGSANALYDOLVTR-----VTHEMAHAGYGN 77
DB 1215 GGARNR--GTLCGPFVTGSRFNFGILRAANDYVITRTNNAENITVNTDITNSPFG 1272
QY 78 GADVGGAGDNSTIELTQNGFRNNATIDOMAKNSDITVGOYGGNNALVNTASDSSVWV 137
DB 1273 ABVGQNV-TRYVNAINTAAVNNLL-----ADSNFVGTIVTISAAL 1319
QY 138 R-----OVFGNNATANQY 151
DB 1320 TTAQDVAKDIAQALGNRLGALRY 1343

RESULT 15
GP63 LEICH STANDARD; PRT; 599 AA.
AC P15706;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Leishmanolysin precursor (EC 3.4.24.36) (Cell surface protease)
DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface
DE endopeptidase).
GN GP63.

```

```

Matches 23; Conservative 5; Mismatches 8; Indels 6; Gaps 2;
QY      52 SAN-AALYDQLVTRVVTHEMAHA-----GYGNGADVGGADN 87
      :|||:|||||||:|:|:|
Db      244 AANIASTRDQLVTRVVTHEMAHALGFSVGFPEGARLIDESISN 285
      :|||:|||||||:|:|:|

Search completed: March 11, 2004, 18:34:59
Job time : 7.3 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:23:54 ; Search time 30.5 Seconds

(without alignments)
1562.074 Million cell updates/sec

Title: US-09-543-407-26

Perfect score: 782

Sequence: 1 MKLKVAFPAALVSGSALA.....DSSVMROYGFGNNTATNOY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL.25.*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	683	87.3	152	2	033802 salmonella
2	580.5	74.2	150	2	07X243 citrobacter
3	532	68.0	149	2	07X240 citrobacter
4	508.5	65.0	152	16	08CW63 escherichia
5	442.5	56.6	150	2	07X237 enterobacte
6	304	38.9	76	2	054069 salmonella
7	122	15.5	29	2	09S3J5
8	121	15.6	502	16	08E1H4
9	114.5	14.6	145	16	08U6N9
10	113	14.5	2174	16	092U08
11	112	14.3	151	16	07UC21
12	112	14.3	160	16	08CW64 escherichia
13	112	14.3	160	16	083RU7
14	110	14.1	151	2	07X244
15	108.5	13.9	1333	16	08BPD3
16	107	13.7	1422	16	08BFD3

17	105.5	13.5	645	16	07U1C5	07U1C5 mycobacteri
18	105.5	13.5	646	16	053818	053818 mycobacteri
19	105	13.4	589	16	07TX53	07TX53 mycobacteri
20	105	13.4	590	16	053309	053309 mycobacteri
21	105	13.4	603	16	08VJ66	08VJ66 mycobacteri
22	104.5	13.4	151	2	07X238	07X238 enterobacte
23	104	13.3	490	16	08EY9	08EY9 leptospira
24	103.5	13.2	91	2	09S3J8	09S3J8 escherichia
25	103.5	13.2	1306	2	093N36	093N36 pantoea ana
26	102.5	13.1	552	16	P96840	P96840 mycobacteri
27	102.5	13.1	552	16	07TW76	07TW76 mycobacteri
28	102.5	13.1	623	16	08V1Y0	08V1Y0 mycobacteri
29	102.5	13.1	1616	2	09KKA1	09KKA1 rickettsia
30	102	13.0	130	16	089J14	089J14 bradyrhizob
31	102	13.0	139	16	08E1H3	08E1H3 shewanella
32	101.5	13.0	152	2	07X241	07X241 citrobacter
33	101.5	13.0	1616	2	09KKA5	09KKA5 rickettsia
34	101	12.9	157	16	08BHG0	08BHG0 pseudomonas
35	100.5	12.9	1616	2	09KKA9	09KKA9 rickettsia
36	100	12.8	191	10	07XDR3	07XDR3 oryza sativ
37	100	12.8	3275	16	08VKM3	08VKM3 mycobacteri
38	100	12.8	3300	16	006304	006304 mycobacteri
39	100	12.8	3507	16	07U270	07U270 mycobacteri
40	99.5	12.7	1613	2	0840U5	0840U5 rickettsia
41	98.5	12.6	154	16	089J15	089J15 bradyrhizob
42	98.5	12.6	453	5	09K6M8	09K6M8 drosophila
43	98.5	12.6	738	5	002402	002402 pinicada fu
44	98.5	12.6	1286	2	0841Y5	0841Y5 campylobact
45	98.5	12.6	3659	16	0981N6	0981N6 rhizobium 1

ALIGNMENTS

RESULT 1	ID	033802	PRELIMINARY:	PRT:	152 AA.
AC	033802;				
DT	01-JAN-1998 (TREMBLrel. 05, Created)				
DT	01-JAN-1998 (TREMBLrel. 05, Last sequence update)				
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)				
DE	Agfa protein (Fragment).				
GN	AGFA.				
OS	Salmonella typhimurium.				
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;				
OC	Enterobacteriaceae; Salmonella.				
OX	NCBI_TaxID=602;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=98053981; PubMed=9393832;				
RA	Sukopolvi S.S., Lorentz R.G., Gordon J.I., Bian Z., Pfeiffer J.D.,				
RA	Normark S.J., Rhen M.;				
RT	"Expression of thin, aggregative fimbriae promotes interaction of				
RT	Salmonella typhimurium SR-11 with mouse small intestinal epithelial				
RT	cells."				
RI	Infect. Immun. 65:5320-5325(1997).				
DR	EMBL, AJ000514; CA04151.1; -				
FT	NON TER 152 152				
SO	SEQUENCE 152 AA; 15401 MW; 9DA7DADC2364B006 CRC64;				
Query Match	87.3%; Score 683; DB 2; Length 152;				
Best Local Similarity	89.4%; Pred. No. 2.8e-48;				
Matches 135; Conservative	3; Mismatches 13; Indels 0; Gaps 0;				
QY	1	MKLKVAFPAALVSGSALAGVPPMGGGGNGGNGSGSPDSTSIYOGSANAALVQ	60		
DB	1	MKLKVAFPAALVSGSALAGVPPMGGGGNGGNGSGSPDSTSIYOGSANAALVQ	60		
QY	61	LVTRVVTHEMAAGYAGNADVQAGDNSTIELTONGFRNNATIDMANKNSDIYGOYGG	120		
DB	61	SPARKSETITTSYGVGNADVQAGDNSTIELTONGFRNNATIDMANKNSDIYGOYGG	120		
QY	121	NNALVNOTASDSSVMROYGFGNNTATNOY 151			

Db 121 NNAALVNQTSDDSVWVROVGFQGNNAIPANQY 151

RESULT 2

Q7X243 PRELIMINARY; PRT; 150 AA.

AC Q7X243;
DT 01-OCT-2003 (TREMBLrel. 25, Created)
DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Culin-csgA protein.
GN CsgA.

OS Citrobacter sp. Fec2.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Citrobacter.
OX NCBI_TaxID=213763;
RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Fec2;

RA Zogaj X., Bokranz W., Nimtz M., Romling U.;

RT "Production of Cellulose and Curli Fimbriae by Members of the Family Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158 (2003).
DR EMBL; AJ515700; CAD5672.1; -

SO SEQUENCE 150 AA; 15016 MW; 1D7141B8D6973DC6 CRC64;

Query Match 74.2%; Score 580.5; DB 2; Length 150;

Best Local Similarity 78.1%; Pred. No. 6.3e-40;

Matches 118; Conservative 10; Mismatches 22; Indels 1; Gaps 1;

QY 1 MKLTKVAFAIIVSGSALAGVPOWGCGGNHNGSGSPDSTLSIYQGSANALAYDQ 60

Db 1 MKLQVAAFAIIVSGSALAGSVPOWGCGGGG-GGGGSSSGPESTLSIYQGVNNAALAQ 59

QY 61 LVTRVVTHEMAHAGYNGADVGGADNSTIELTONGFRNNAITIDONNAKSDITVGGY 120

Db 60 SDARKSDVTTHQNGRNGADVGGSDNSTIDLTONGFRNNAITIDONNAKSDITVGGY 119

QY 121 NNAALVNQTSDDSVWVROVGFQGNNAIPANQY 151

Db 120 HNAALVNQTSDDSVLVHVGFGNNAIPANQY 150

RESULT 3

Q7X240 PRELIMINARY; PRT; 149 AA.

AC Q7X240;
DT 01-OCT-2003 (TREMBLrel. 25, Created)

DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Culin-csgA protein.
GN CsgA.

OS Citrobacter freundii.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Citrobacter.

OX NCBI_TaxID=546;
RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Fec4;

RA Zogaj X., Bokranz W., Nimtz M., Romling U.;

RT "Production of Cellulose and Curli Fimbriae by Members of the Family Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158 (2003).
DR EMBL; AJ515701; CAD5675.1; -

SO SEQUENCE 149 AA; 15260 MW; 946D52017F648FD CRC64;

Query Match 68.0%; Score 532; DB 2; Length 149;

Best Local Similarity 71.5%; Pred. No. 5.7e-36;

Matches 108; Conservative 17; Mismatches 24; Indels 2; Gaps 1;

QY 1 MKLTKVAFAIIVSGSALAGVPOWGCGGNHNGSGSPDSTLSIYQGSANALAYDQ 60

Db 1 MKLTKVAFAIIVSGSALAGVPOW--GNNHGGGSGNYGPDSSLSIYQGSNNANALAQ 58

QY 61 LVTRVVTHEMAHAGYNGADVGGADNSTIELTONGFRNNAITIDONNAKSDITVGGY 120

Db 59 SDARKSDVTTHQNGRNGADVGGADNSTIELTONGFRNNAITIDONNAKSDITVGGY 118

QY 121 NNAALVNQTSDDSVWVROVGFQGNNAIPANQY 151

Db 119 RNCALVNQTSDDSVNLVQVGFQGNNAIPANQY 149

RESULT 4

Q8CW63 PRELIMINARY; PRT; 152 AA.

AC Q8CW63;
DT 01-MAR-2003 (TREMBLrel. 23, Created)

DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Major curlin subunit precursor.
GN CsgA OR C1306.

OS Escherichia coli O6.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Escherichia.

OX NCBI_TaxID=217992;
RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=O6.H1 / ATCC 700928;

RX MEDLINE=22388234; PubMed=12471157;

RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,

RA Rasbo D., Buckles E.L., Ilou S.-R., Boutin A., Hackett J., Stroud D.,

RA Mobley G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,

RT "Extensive mosaic structure revealed by the complete genome sequence

of uropathogenic Escherichia coli.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).

DR EMBL; AE016759; AAN7979.1; -

KW Complete proteome.

SO SEQUENCE 152 AA; 15064 MW; 3BA57P34C1240E83 CRC64;

Query Match 65.0%; Score 508.5; DB 16; Length 152;

Best Local Similarity 67.8%; Pred. No. 4.8e-34;

Matches 103; Conservative 18; Mismatches 30; Indels 1; Gaps 1;

QY 1 MKLTKVAFAIIVSGSALAGVPOW--GCGGNHNGSGNSGPDSTLSIYQGSANALAYD 59

Db 1 MKLTKVAFAIIVSGSALAGVPOYGGGNGHGGGNSGPNSELTITVGGNSALAQ 60

QY 60 QLVTRVVTHEMAHAGYNGADVGGADNSTIELTONGFRNNAITIDONNAKSDITVGGY 119

Db 61 QADARNSDLTTHQNGRNGADVGGSDNSTIDLTORGFSNATIDONNAKSDITVGGY 120

QY 120 GNNALVNQTSDDSVWVROVGFQGNNAIPANQY 151

Db 121 GNGALVNDQTSNDSVNVTVGFQGNNAIPANQY 152

RESULT 5

Q7X237 PRELIMINARY; PRT; 150 AA.

AC Q7X237;
DT 01-OCT-2003 (TREMBLrel. 25, Created)

DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Culin-csgA protein.
GN CsgA.

OS Enterobacter sakazakii.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Enterobacter.

OX NCBI_TaxID=28141;
RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Fec39;

RA Zogaj X., Bokranz W., Nimtz M., Romling U.;


```

Q8U6N9          PRELIMINARY;          PRT;          145 AA.
ID Q8U6N9
AC Q8U6N9;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein Adu4768.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
NCBI_TaxID=176299;
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
Oura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
Chapman P., Clendinning J., Deatherage G., Gillet W., Grant C.,
Kutyavin T., Levy R., Li M.-J., McCelland E., Palmeri A., Gordon D.,
Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
Nester E.W.;
RA "The genome of the natural genetic engineer Agrobacterium tumefaciens
C58."
RT Science 294:2317-2323(2001).
RL [2]
RN SEQUENCE FROM N.A.
RP MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
Houmel K., Gordon J., Vaudin M., Iarchouk O., Epp A., Liu F.,
Wollam C., Allinger M., Doughy D., Scott C., Lappas C., Marfelz B.,
Rangan C., Crowell C., Gerson J., Lomo C., Sear C., Strub G.,
Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
Agrobacterium tumefaciens C58."
RL Science 294:2323-2328(2001).
DR EMBL; AE009405; AAL45562.1; -.
DR EMBL; AE008209; AAK86882.1; -.
DR PIR; AD3143; AD3143.
DR PIR; H98144; H98144.
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 145 AA; 14984 MW; DEDC870E1713D51A CRC64;

Query Match          14.6%; Score 114.5; DB 16; Length 145;
Best Local Similarity 25.2%; Pred. No. 0.064;
Matches 39; Conservative 26; Mismatches 59; Indels 31; Gaps 4;

QY 3 LKVAFAALVVGSGALAGVVPOMG-----GNHNGGNSGSPDSTLSIYXGANA 55
DB 1 MRKSFIALVALVGLSAAAPMANADVRIEYQGSNSAGQEGGNRIPTYQNGYV- 59
QY 56 ALYDQVTRVVTHEMAHAGVNGADYVGADNSTIELTNGFRNNTTIDQWAKKSDITV 115
DB 60 -----RIVGHQY---GRHLSAVGQEGHNDYSGTTQNGKRVNAGI----- 96
QY 116 GQYGNNAALVNOTASDSSVWVRQVFGNNATANO 150
DB 97 GQFGSNHTTILTDQNGNIAAGVQVGRGGSANVSQ 131

RESULT 10
Q92U08          PRELIMINARY;          PRT;          2174 AA.
ID Q92U08;
AC Q92U08;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical glycine-rich protein SMD21548.
GN RB0989 OR SMD21548.

```

```

OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Plasmid pSymB (megaplasmid 2).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
NCBI_TaxID=382;
RX MEDLINE=21396508; PubMed=11481431;
RA Fian T.M., Weidner S., Wong K., Buhrmester J., Chain P.,
Vorholter F.J., Hernandez-Lucas I., Becker A., Cowie A., Gouy J.,
Golding B., Puhler A.;
RT "The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-
fixing endosymbiont Sinorhizobium meliloti."
Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).
DR EMBL; AL603645; CAC49389.1; -.
DR PIR; B95965; B95965.
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
DR GO; GO:0004601; F:peroxidase activity; IEA.
DR GO; GO:0006979; P:response to oxidative stress; IEA.
DR InterPro; IPR005546; Autotransporter.
DR InterPro; IPR002016; Peroxidase.
DR InterPro; IPR002173; PEK.
DR Pfam; PF03797; Autotransporter; 1.
DR PROSITE; PS00435; PROXIDASE_1; 1.
DR PROSITE; PS00583; PEK_KINASES_1; 2.
DR Plasmid; Hypothetical protein; Complete proteome.
SQ SEQUENCE 2174 AA; 203314 MW; 008EB68297B44182 CRC64;

Query Match          14.5%; Score 113; DB 16; Length 2174;
Best Local Similarity 27.0%; Pred. No. 2.3;
Matches 40; Conservative 21; Mismatches 51; Indels 36; Gaps 7;

QY 11 AIVVSGALAGVVPQ--WGCGNHNGGNSGSPDSTLSIYXGANAALYDQVTRVTH 68
DB 693 AIAVAGAGAVGIIAQSIGGGGN---GQNTGDDAGRGSPQIGGGGG----- 737
QY 69 EMAHAGVGNADVG-----QGANSTI--ELTNGFRNNTTIDQWAKKSDITV 115
DB 738 ---GGYANTANVAFKGLTTLTQGSAAAGIVAQVGGGGGTGTGASYSIGIGFTASVAV 793
QY 116 GQYGNNAALVNOTASDSSVWVRQV 141
DB 794 GGTGNGAGAGGVSVSLTDSAIRTGCG 821

RESULT 11
Q7UCZ1          PRELIMINARY;          PRT;          151 AA.
ID Q7UCZ1
AC Q7UCZ1;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Minor curlin subunit.
GN CSGB OR S1108.
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Shigella.
NCBI_TaxID=623;
RX MEDLINE=24557T / ATCC 700930 / serotype 2a;
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
Schwartz D.C., Blattner F.R.;
RT "Complete genome sequence and comparative genomics of Shigella
flexneri serotype 2a strain 2457T."
Intec. Immun. 71:2775-2786(2003).
DR EMBL; AE016981; AAP16542.1; -.
SQ SEQUENCE 151 AA; 15868 MW; 5D5D26B964014A0 CRC64;

```


RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 33913 / NCPPB 528;
 RX MEDLINE=2022145; PubMed=12024217;
 RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
 RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
 RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.B.A.,
 RA Camarotte G., Camavon F., Cardozo J., Chamberg F., Ciapina L.P.,
 RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorxy H.,
 RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
 RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
 RA Katayama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
 RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
 RA Martins E.C., Melandis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
 RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
 RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
 RA Spínola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezsa R.I.D.,
 RA Trindade dos Santos M., Truffi D., Tsai S.M., White P.F.,
 RA Setubal J.C., Kitajima J.P.,
 RT "Comparison of the genomes of two Xanthomonas pathogens with differing
 RT host specificities."
 RL Nature 417:459-463(2002).
 RL EMBL; AE012148; AAM39823.1; -
 DR GO; GO:0009279; C:external outer membrane (sensu Gram-negativ. . .; IEA.
 DR InterPro; IPR000258; Ice_nucleatn.
 DR Pfam; PF00818; Ice_nucleation; 68.
 DR PRINTS; PR00327; ICNUCLEATN.
 DR PROSITE; PS00314; ICE_NUCLEATION; 40.
 DR Complete proteome.
 SQ SEQUENCE 1333 AA; 131306 MW; 326078458D0E4842 CRC64;

Query Match 13.9%; Score 108.5; DB 16; Length 1333;
 Best Local Similarity 26.3%; Pred. No. 3;
 Matches 46; Conservative 22; Mismatches 48; Indels 59; Gaps 9;

QY 5 KVAAPAAIVSG-----SALAGVPPQWGGGNGGNSGPDSTLSIYQYGA 53
 DB 856 QTAGYKSLITGYGSGTQTAQESSSLIAGY-----GSSMAGPDSL-IAGYGT 903
 QY 54 NAALYDQVTRVVTHEMAHAGYGCADVGCGADNSTELTQNGFRNATID----- 104
 DB 904 QTAGYDSFLT-----AGYGS---TQTAQSSWLITGYGSTASTASFOSSLIAGYGS 950
 QY 105 -QWNAKNSDITVQGYGNNAL-----VNQTASDSSVMVRYQVFGNNATA 148
 DB 951 TQTAGYESTLTAG-YGSTQTAQEISWLTGYGSTQTAGHGSILT--AGYGSNSTA 1002

Search completed: March 11, 2004, 18:40:46
 Job time : 31.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:13:53 ; Search time 45.9 Seconds
(without alignments)
929.514 Million cell updates/sec

Title: US-09-543-407-28

Perfect score: 775
Sequence: 1 MGLIKVAAFAIVSGSALA.....DSVMVRYGFQNNATANY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database: A_Geneseq_29Jan04:*

1: geneseqp19808:.*
2: geneseqp19908:.*
3: geneseqp20008:.*
4: geneseqp20018:.*
5: geneseqp20028:.*
6: geneseqp20038:.*
7: geneseqp20038:.*
8: geneseqp20048:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	775	100.0	151	3	AAB36354 Agfa::PT3
2	712	91.9	151	3	AAB36351 Agfa::PT3
3	683	88.1	151	2	AAB74625 Agfa sequ
4	683	88.1	151	3	AAB36341 Salmonell
5	678	87.5	151	2	AAB23570 Salmonell
6	662	85.4	151	3	AAB36350 Agfa::PT3
7	617	79.6	151	3	AAB36353 Agfa::PT3
8	613	79.1	151	3	AAB36355 Agfa::PT3
9	605	78.1	151	3	AAB36346 Agfa::PT3
10	603	77.8	151	3	AAB36347 Agfa::PT3
11	601	77.5	151	3	AAB36349 Agfa::PT3
12	600	77.4	151	3	AAB36352 Agfa::PT3
13	568	73.3	151	3	AAB36348 Agfa::PT3
14	509	65.7	151	3	AAB36343 Escherich
15	504	65.0	151	7	ABR82651 E. coli C
16	498	64.3	120	2	AAB2761 Agfa sequ
17	488	64.3	120	2	AAB23569 Salmonell
18	435	56.1	142	2	AAB52664 Fibronect
19	359	46.3	122	2	AAB52663 FMB curli
20	147	19.0	45	3	AAB52636 Salmonell
21	132	17.0	22	3	AAB36318 Salmonell
22	113	14.6	24	7	ABR82644 E. coli c
23	111	14.3	22	3	AAB36332 Salmonell
24	111	14.3	22	3	AAB36327 Salmonell
25	111	14.3	22	3	AAB36337 Salmonell

26	109	14.1	23	3	AAB36340	AAB36340 Salmonell
27	109	14.1	23	3	AAB36324	AAB36324 Salmonell
28	109	14.1	23	3	AAB36319	AAB36319 Salmonell
29	102	13.2	26	7	ABR82649	Abt82649 E. coli V
30	96	12.4	19	3	AAB36323	AAB36323 Salmonell
31	96	12.4	19	3	AAB36336	AAB36336 Salmonell
32	96	12.4	19	3	AAB36328	AAB36328 Salmonell
33	94.5	12.2	151	3	AAB36342	AAB36342 Salmonell
34	92	11.9	1074	6	ABU22692	Abu22692 Protein e
35	91.5	11.8	502	2	AAB32312	Aab32312 Leishman
36	90	11.6	24	7	ABR82642	Abt82642 E. coli N
37	89.5	11.5	151	3	AAB36344	Abt82644 Escherich
38	89	11.5	24	7	ABR82647	Abt82647 E. coli C
39	89	11.5	354	7	ABO23520	ABO23520 Mycodacte
40	87.5	11.3	249	3	AAY69523	Aay69523 Ant1-CD38
41	87.5	11.3	447	3	AA297278	Aag297278 Arabidops
42	87.5	11.3	468	3	AA297277	Aag297277 Arabidops
43	87.5	11.3	842	5	ABP66189	Abp66189 Bifidobac
44	87.5	11.3	1419	5	ABP69842	Abp69842 Human pol
45	87.5	11.3	1477	5	ABP69841	Abp69841 Human pol

ALIGNMENTS

RESULT 1	AAB36354	standard; protein; 151 AA.
ID	AAB36354	
XX		
AC	AAB36354;	
XX		
DT	26-FEB-2001 (first entry)	
DE	Agfa::PT3#9 amino acid sequence SEQ ID NO:28.	
XX		
KW	Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;	
KW	vaccine; immune response; immunogen.	
XX		
OS	Salmonella enteritidis.	
OS	Escherichia coli.	
OS	Synthetic.	
XX		
FN	WO200060102-A2.	
XX		
PD	12-OCT-2000.	
XX		
PF	05-APR-2000; 2000WO-CA000356.	
XX		
PR	05-APR-1999; 99US-0127888P.	
XX		
PA	(UYVI-) UNIV VICTORIA.	
PI	White AP, Doran JL, COLLISON SK, Kay MW;	
XX		
DR	WPI; 2000-672631/65.	
XX		
NR	N-PSDB; AAC64630.	
PT	Recombinant agfa gene having a segment replaced by a foreign DNA sequence	
PT	which encodes foreign epitope or antigen, expresses recombinant Agfa	
PT	protein useful for eliciting immune response in animal.	
XX		
PS	Disclosure; Page 138, 139pp; English.	
XX		
CC	The present invention describes a recombinant agfa gene (1) where a	
CC	segment of the gene has been replaced by a segment of a foreign DNA	
CC	sequence which encodes a foreign epitope or antigen. Also described are:	
CC	(1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended	
CC	assembly system of strains of Salmonella, Escherichia coli and	
CC	enterobacteriaceae for the production of fimbriae comprising recombinant	
CC	Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)	
CC	directing recombination of a recombinant gene into the chromosome of the	
CC	homologous species; (3) directing recombination of a recombinant gene	
CC	back into the chromosome of the homologous species, replacing the native	

CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fibrillar presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell) the hybrid fibrin protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fibrillar subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fibrillae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

Sequence 151 AA;

Query Match 100.0%; Score 775; DB 3; Length 151;
 Best Local Similarity 100.0%; Pred. No. 2.5e-67;
 Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLKVAAFAAIVVSGSALAGVVPWGGGNNHNGGNSGPDSTLSTIYGSANAALALQ 60
 DB 1 MKLKVAAFAAIVVSGSALAGVVPWGGGNNHNGGNSGPDSTLSTIYGSANAALALQ 60

QY 61 SDARKSETTITOSGNGADYDQLVTRVTHEMAHAFNNNAITIDQMNKNSDITVGOYGG 120
 DB 61 SDARKSETTITOSGNGADYDQLVTRVTHEMAHAFNNNAITIDQMNKNSDITVGOYGG 120

QY 121 NNAALVNOTASDSSVMVQVGFNNATANOY 151
 DB 121 NNAALVNOTASDSSVMVQVGFNNATANOY 151

RESULT 2
 AAB36351
 ID AAB36351 standard; protein; 151 AA.
 AC AAB36351;
 XX
 DT 26-FEB-2001 (first entry)
 DE Agfa::PT3#6 amino acid sequence SEQ ID NO:22.
 KW *Salmonella*; agfa; chromosomal gene replacement; fibrin; epitope;
 XX vaccine; immune response; immunogen.
 OS *Salmonella enteritidis*.
 OS *Escherichia coli*.
 OS Synthetic.
 XX
 PN WO200060102-A2.
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay WW;
 DR WPI: 2000-672631/65.
 DR N-PSDB; AAC64627.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 137; 139pp; English.

XX The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fibrillae (SEF17/7AF) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Enterobacteriaceae* for the production of fibrillae comprising recombinant
 CC Agfa, Caga and Agfa-homologue fibrin subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fibrillar presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell) the hybrid fibrin protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fibrillar subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fibrillae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

Sequence 151 AA;

Query Match 91.9%; Score 712; DB 3; Length 151;
 Best Local Similarity 91.1%; Pred. No. 3.3e-61;
 Matches 144; Conservative 0; Mismatches 0; Indels 14; Gaps 2;

QY 1 MKLKVAAFAAIVVSGSALAGVVPWGGGNNHNGGNSGPDSTLSTIYGSANAALALQ 60
 DB 1 MKLKVAAFAAIVVSGSALAGVVPWGGGNNHNGGNSGPDSTLSTIYGSANAALALQ 60

QY 61 SDARKSETTITOSGNGADYDQLVTRVTHEMAHAFNNNAITIDQMNKNSDI 113
 DB 61 SDARKSETTITOSGNGADYDQLVTRVTHEMAHAFNNNAITIDQMNKNSDI 113

QY 114 TVGOYGGNNAALVNOTASDSSVMVQVGFNNATANOY 151
 DB 114 TVGOYGGNNAALVNOTASDSSVMVQVGFNNATANOY 151

RESULT 3
 AAR74625
 ID AAR74625 standard; protein; 151 AA.
 AC AAR74625;
 XX
 DT 25-MAR-2003 (revised)
 DT 26-JUN-1995 (first entry)
 DE Agfa sequence.
 KW *Salmonella*; Agfa; vaccine.
 XX *Salmonella*.
 OS *Salmonella*.
 OS WO9425598-A2.
 PN WO9425598-A2.
 PD 10-NOV-1994.
 XX
 PF 26-APR-1994; 94WO-IB000207.
 XX
 PR 26-APR-1993; 93US-00054452.
 XX
 PA (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
 PA (KING/) KING J.
 XX

PI Kay MW, Collinson SK, Clouthier SC, Doran JL;
 XX WPI, 1994-358275/44.
 DR N-PSDB; AAC67467.
 XX
 PT Eliciting an immune response to *Salmonella* - using attenuated *Salmonella*
 PT strains, vector constructs, or compens. contg. fimbrial type proteins.
 XX
 PS Disclosure; Fig 7B; 95pp; English.
 XX
 CC The *Salmonella* Agfa protein and DNA are used in vaccine and genetic
 CC immunization compositions, respectively, to elicit an immune response to
 CC *Salmonella* in animals (e.g. food producing animals) and humans. (Updated
 CC on 25-MAR-2003 to correct PN field.)
 XX
 SQ Sequence 151 AA;
 Query Match 88.1%; Score 683; DB 2; Length 151;
 Best Local Similarity 90.1%; Pred. No. 2.2e-58;
 Matches 136; Conservative 1; Mismatches 14; Indels 0; Gaps 0;
 QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYQGSANALALQ 60
 DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYQGSANALALQ 60
 QY 61 SDAKSETTITQSGYNGADYDQLVTRVVTHEMAHFRNNATIDQNNAKSDITVQYGG 120
 DB 61 SDAKSETTITQSGYNGADYDQLVTRVVTHEMAHFRNNATIDQNNAKSDITVQYGG 120
 QY 121 NNAALVNOTASDSSVWVRQVFGNNATANQY 151
 DB 121 NNAALVNOTASDSSVWVRQVFGNNATANQY 151
 RESULT 4
 AAB36341
 ID AAB36341 standard; protein; 151 AA.
 XX
 AC AAB36341;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE *Salmonella* enteritidis Agfa amino acid sequence SEQ ID NO:5.
 XX
 KW *Salmonella*; agfa; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS *Salmonella* enteritidis.
 XX
 PN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000MO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collinson SK, Kay MW;
 XX
 DR WPI; 2000-672631/65.
 DR N-PSDB; AAC64617.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 135; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:

CC (1) use of thin aggregative fimbriae (SEPT17/TAI) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 SQ Sequence 151 AA;
 Query Match 88.1%; Score 683; DB 3; Length 151;
 Best Local Similarity 90.1%; Pred. No. 2.2e-58;
 Matches 136; Conservative 1; Mismatches 14; Indels 0; Gaps 0;
 QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYQGSANALALQ 60
 DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYQGSANALALQ 60
 QY 61 SDAKSETTITQSGYNGADYDQLVTRVVTHEMAHFRNNATIDQNNAKSDITVQYGG 120
 DB 61 SDAKSETTITQSGYNGADYDQLVTRVVTHEMAHFRNNATIDQNNAKSDITVQYGG 120
 QY 121 NNAALVNOTASDSSVWVRQVFGNNATANQY 151
 DB 121 NNAALVNOTASDSSVWVRQVFGNNATANQY 151
 RESULT 5
 AAW23570
 ID AAW23570 standard; protein; 151 AA.
 XX
 AC AAW23570;
 XX
 DT 25-MAR-2003 (revised)
 DT 29-SEP-1997 (first entry)
 XX
 DE *Salmonella* enteritidis 27655-3b agfa.
 XX
 KW Enteropathogenic bacteria; enterobacteria; *S. enteritidis*; antibody.
 XX
 OS *Salmonella* enteritidis.
 XX
 PI Key Location/Qualifiers
 FT Misc-difference /note= "Encoded by GCC"
 FT
 XX
 PN US5635617-A.
 XX
 PD 03-JUN-1997.
 XX
 PF 26-APR-1994; 94US-00233788.
 XX
 PR 26-APR-1993; 93US-00054452.
 XX
 PA (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
 XX
 PI Collinson SK, Kay MW, Doran JL;

XX WPI; 1997-309886/28.
 DR N-PSDB; AAT74142.
 XX Isolated *Salmonella* gene agfa - used for diagnosis of *Salmonella* or
 PT enteropathogenic bacteria of the Enterobacteria family.
 XX Example 2; Fig 7; 85pp; English.
 XX The present sequence represents agfa encoded by the full agfa gene
 CC derived from *Salmonella enteritidis* 27655-3b. The nucleic acid can be
 CC used to provide diagnostic assays for *Salmonella* and/or enteropathogenic
 CC bacteria of the family Enterobacteria. It can also be used to provide
 CC proteins and antibodies which can be used for assays. The nucleic acid
 CC sequence can be used to provide probes or primers which can specifically
 CC hybridise to nucleic acid molecules from greater than 99% of *Salmonella*
 CC strains that are pathogenic to warm-blooded animals relative to nucleic
 CC acid molecules from virtually all other microbial organisms. (Updated on
 CC 25-MAR-2003 to correct PF field.)
 XX Sequence 151 AA;
 SQ
 Query Match 87.5%; Score 678; DB 2; Length 151;
 Best Local Similarity 89.4%; Pred. No. 6.6e-58;
 Matches 135; Conservative 1; Mismatches 15; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAFAIIVSGSALAGVVPQWGGGNNHGGSSGPDSTLSTIYQGSANAALALQ 60
 DB 1 MKLLKVAFAFAIIVSGSALAGVVPQWGGGNNHGGSSGPDSTLSTIYQGSANAALALQ 60
 QY 61 SDARKSETTITQSGYGNAGADYDQLVTRVVTHEMAHAFRNATIDQNAKNSDITVGQYGG 120
 DB 61 SDARKSETTITQSGYGNAGADYDQLVTRVVTHEMAHAFRNATIDQNAKNSDITVGQYGG 120
 QY 121 NNAALVNOTASDSSVMVROVGFNNATATNOY 151
 DB 121 NNPALVNOTASDSSVMVROVGFNNATATNOY 151
 RESULT 6
 AAB36350
 ID AAB36350 standard; protein; 151 AA.
 AC AAB36350;
 XX 26-FEB-2001 (first entry)
 DT
 XX Agfa::PT3#5 amino acid sequence SEQ ID NO:20.
 DE
 XX *Salmonella*; agfa; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 XX *Salmonella enteritidis*.
 OS *Escherichia coli*.
 OS Synthetic.
 OS
 XX WO200060102-A2.
 FN
 XX 12-OCT-2000.
 PD
 XX 05-APR-2000; 2000WO-CA000356.
 PF
 XX 05-APR-1999; 99US-0127888P.
 PR
 XX (UYVI-) UNIV VICTORIA.
 PA
 XX White AP, Doran JL, Collison SK, Kay WW,
 PI
 XX WPI; 2000-672631/65.
 DR
 XX N-PSDB; AAC64626.
 DR
 XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa

PT protein useful for eliciting immune response in animal.
 XX
 XX Disclosure; Page 137; 139pp; English.
 PS
 XX The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF7/7AF) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, GsgA and Agfa-homologue fimbria subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX Sequence 151 AA;
 SQ
 Query Match 85.4%; Score 662; DB 3; Length 151;
 Best Local Similarity 81.9%; Pred. No. 2.4e-56;
 Matches 136; Conservative 0; Mismatches 0; Indels 30; Gaps 2;
 QY 1 MKLLKVAFAFAIIVSGSALAGVVPQWGGGNNHGGSSGPDSTLSTIYQGSANAALALQ 60
 DB 1 MKLLKVAFAFAIIVSGSALAGVVPQWGGGNNHGGSSGPDSTLSTIYQGSANAALALQ 60
 QY 61 SDARKSETTITQSGYGNAGADYDQLVTRVVTHEMAHAFRNATIDQ 105
 DB 61 SDARKSETTITQSGYGNAGADYDQLVTRVVTHEMAHAFRNATIDQ 105
 QY 106 WNAKNSDITVGQYGNALVNOTASDSSVMVROVGFNNATATNOY 151
 DB 106 WNAKNSDITVGQYGNALVNOTASDSSVMVROVGFNNATATNOY 151
 RESULT 7
 AAB36353
 ID AAB36353 standard; protein; 151 AA.
 AC AAB36353;
 XX 26-FEB-2001 (first entry)
 DT
 XX Agfa::PT3#8 amino acid sequence SEQ ID NO:26.
 DE
 XX *Salmonella*; agfa; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 XX *Salmonella enteritidis*.
 OS *Escherichia coli*.
 OS Synthetic.
 OS
 XX WO200060102-A2.
 FN
 XX 12-OCT-2000.
 PD
 XX 05-APR-2000; 2000WO-CA000356.
 PF

PR 05-APR-1999; 99US-0127888P.
 XX
 XX (UYVI-) UNIV VICTORIA.
 PA White AP, Doran JL, Collison SK, Kay WW,
 PI MPI; 2000-672631/65.
 XX DR N-PSDB; AAC64629.
 DR
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 XX protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 138; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/7AP) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 XX Sequence 151 AA:
 S0
 Query Match 79.6%; Score 617; DB 3; Length 151;
 Best Local Similarity 73.6%; Pred. No. 5.7e-52;
 Matches 128; Conservative 0; Mismatches 0; Indels 46; Gaps 2;
 QY 1 MKLLKVAAPAAIVVSGSALAGVVPWGGGNNHNGGNSGPDSTLSIYOGSNNALALQ 60
 DB 1 MKLLKVAAPAAIVVSGSALAGVVPWGGGNNHNGGNSGPDSTLSIYOGSNNALALQ 57
 QY 61 SDAKSETTITGSGNGADYDQLVTRVVTHEMAHMA-----F 97
 DB 58 -----YDQLVTRVVTHEMAHMAHGYGNGADVGGADNSTELTQNGF 97
 QY 98 RNNATIDOWNNAKNSDITVGGYGNNALVNOTASDSSVWVRQVFGNNATANQY 151
 DB 98 RNNATIDOWNNAKNSDITVGGYGNNALVNOTASDSSVWVRQVFGNNATANQY 151
 RESULT 8
 AAB36355
 ID AAB36355 standard; protein; 151 AA.
 XX
 AC AAB36355;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Agfa::PT3#10 amino acid sequence SEQ ID NO:30.
 XX
 KW *Salmonella*; agfa; chromosomal gene replacement; fimbrial; epitope;
 KW vaccine; immune response; immunogen.

XX
 XX *Salmonella enteritidis*.
 OS *Escherichia coli*.
 OS Synthetic.
 XX
 PN MO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 XX 05-APR-2000; 2000WO-CN000356.
 PF
 XX
 XX 05-APR-1999; 99US-0127888P.
 PR
 XX (UYVI-) UNIV VICTORIA.
 PA White AP, Doran JL, Collison SK, Kay WW,
 PI MPI; 2000-672631/65.
 XX DR N-PSDB; AAC64631.
 DR
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 XX protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 139; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/7AP) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 XX Sequence 151 AA:
 S0
 Query Match 79.1%; Score 613; DB 3; Length 151;
 Best Local Similarity 74.6%; Pred. No. 1.4e-51;
 Matches 129; Conservative 0; Mismatches 0; Indels 44; Gaps 2;
 QY 1 MKLLKVAAPAAIVVSGSALAGVVPWGGGNNHNGGNSGPDSTLSIYOGSNNALALQ 60
 DB 1 MKLLKVAAPAAIVVSGSALAGVVPWGGGNNHNGGNSGPDSTLSIYOGSNNALALQ 60
 QY 61 SDAKSETTITGSGNGADYDQLVTRVVTHEMAHMA-----YDQLVTRVVTHEMAHMA 98
 DB 61 SDAKSETTITGSGNGADVGGADNSTELTQNGFRNNATVTDQLVTRVVTHEMAHMA-- 118
 QY 99 RNNATIDOWNNAKNSDITVGGYGNNALVNOTASDSSVWVRQVFGNNATANQY 151
 DB 119 -----GGNNALVNOTASDSSVWVRQVFGNNATANQY 151
 RESULT 9

AAB36346
ID AAB36346 standard; protein; 151 AA.
XX
AC AAB36346;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa::PT3#1 amino acid sequence SEQ ID NO:12.
XX
KW Salmomella; agfa; chromosomal gene replacement; fimbria; epitope;
KW vaccine; immune response; immunogen.
XX
OS Salmomella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO20060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI: 2000-672631/65.
DR N-PSDB; AAC64622.
XX
PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 135; 139pp; English.
XX
CC The present invention describes a recombinant agfa gene (1) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SFF17/TAF) nucleation depended
CC assembly system of strains of Salmomella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmomella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 151 AA:
Query Match 78.1%; Score 605; DB 3; Length 151;
Best Local Similarity 80.1%; Pred. No. 8.3e-51;
Matches 121; Conservative 5; Mismatches 25; Indels 0; Gaps 0;
QY 1 MKLIKVAAPAIIVSGSALAGVVPOMGGGNNHNGGSSGPDSTLSTIYOGSANAALATQ 60
DB 1 MKLIKVAAPAIIVSGSALAGVVPOMGGGNNHNGGSSGPDSTLSTIYOGSANAALATQ 60

QY 61 SDARKSETTITGSGYNGADYDQLVTRVVTHEMAHAFRNATIDQNNAKSDITVGQYGG 120
DB 61 SDARKSETTITGSGYNGADYDQLVTRVVTHEMAHAFRNATIDQNNAKSDITVGQYGG 120
QY 121 NNAALVNDQASDSSVMRVQFGNNATANQY 151
DB 121 NNAALVNDQLVTRVVTHEMAHANNATANQY 151
RESULT 10
AAB36347
ID AAB36347 standard; protein; 151 AA.
XX
AC AAB36347;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa::PT3#2 amino acid sequence SEQ ID NO:14.
XX
KW Salmomella; agfa; chromosomal gene replacement; fimbria; epitope;
KW vaccine; immune response; immunogen.
XX
OS Salmomella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO20060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI: 2000-672631/65.
DR N-PSDB; AAC64623.
XX
PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 136; 139pp; English.
XX
CC The present invention describes a recombinant agfa gene (1) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SFF17/TAF) nucleation depended
CC assembly system of strains of Salmomella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmomella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in

CC the exemplification of the present invention
XX Sequence 151 AA;
SQ

Query Match 77.8%; Score 603; DB 3; Length 151;
Best Local Similarity 80.8%; Pred. No. 1.3e-50;
Matches 122; Conservative 4; Mismatches 25; Indels 0; Gaps 0;

QY 1 MKLKYAAPAAIYVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYOGSANAALALQ 60
DB 1 MKLKYAAPAAIYVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYOGSANAALALQ 60
QY 61 SDARKSETTTTOSGYNGADYDQLVTRVVTHEMAHAFRNNAATTDQNAKNSDITVGOYGG 120
DB 61 SDARKSETTTTOSGYNGADYDQLVTRVVTHEMAHAFRNNAATTDQNAKNSDITVGOYGG 120
QY 121 NNAALVNOTASDSSVWVROYGFGNNATANQY 151
DB 121 LVTTRVTHEMAHAFRNNAATTDQNAKNSDITVGOYGG 151

RESULT 11
AAB36349
ID AAB36349 standard; protein; 151 AA.
AC AAB36349;
XX
XX
XX 26-FEB-2001 (first entry)
DE Agfa::PT3#4 amino acid sequence SEQ ID NO:18.
XX
XX
XX Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
KW vaccine; immune response; immunogen.
XX
XX OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
XX WO200060102-A2.
XX
XX PD 12-OCT-2000.
XX
XX PF 05-APR-2000; 2000WO-CA000356.
XX
XX PR 05-APR-1999; 99US-0127888P.
XX
XX PA (UYVI-) UNIV VICTORIA.
XX
XX PI White AP, Doran JL, Collison SK, Kay WW;
XX
XX WPI; 2000-672631/65.
XX
XX DR N-PSDB; AAC64625.
XX
XX PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
XX PT which encodes foreign epitope or antigen, expresses recombinant Agfa
XX PT protein useful for eliciting immune response in animal.
XX
XX PS Disclosure; Page 136; 139pp; English.

CC The present invention describes a recombinant agfa gene (1) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAI) nucleation depended
CC assembly system of strains of salmonella, Escherichia coli; and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal.
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or

CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention

QY Sequence 151 AA;
SQ

Query Match 77.5%; Score 601; DB 3; Length 151;
Best Local Similarity 80.8%; Pred. No. 2e-50;
Matches 122; Conservative 4; Mismatches 25; Indels 0; Gaps 0;

QY 1 MKLKYAAPAAIYVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYOGSANAALALQ 60
DB 1 MKLKYAAPAAIYVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYOGSANAALALQ 60
QY 61 SDARKSETTTTOSGYNGADYDQLVTRVVTHEMAHAFRNNAATTDQNAKNSDITVGOYGG 120
DB 61 SDARKSETTTTOSGYNGADYDQLVTRVVTHEMAHAFRNNAATTDQNAKNSDITVGOYGG 120
QY 121 NNAALVNOTASDSSVWVROYGFGNNATANQY 151
DB 121 NNAALVNOTASDSSVWVROYGFGNNATANQY 151

RESULT 12
AAB36352
ID AAB36352 standard; protein; 151 AA.
AC AAB36352;
XX
XX
XX 26-FEB-2001 (first entry)
DE Agfa::PT3#7 amino acid sequence SEQ ID NO:24.
XX
XX
XX
KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
KW vaccine; immune response; immunogen.
XX
XX OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
XX PN WO200060102-A2.
XX
XX PD 12-OCT-2000.
XX
XX PF 05-APR-2000; 2000WO-CA000356.
XX
XX PR 05-APR-1999; 99US-0127888P.
XX
XX XX
XX PA (UYVI-) UNIV VICTORIA.
XX
XX PI White AP, Doran JL, Collison SK, Kay WW;
XX
XX WPI; 2000-672631/65.
XX
XX DR N-PSDB; AAC64628.
XX
XX PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
XX PT which encodes foreign epitope or antigen, expresses recombinant Agfa
XX PT protein useful for eliciting immune response in animal.
XX
XX PS Disclosure; Page 138; 139pp; English.

CC The present invention describes a recombinant agfa gene (1) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:

CC (1) use of thin aggregative fimbriae (SEF17/TAf) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

CC Sequence 151 AA;

Query Match 77.4%; Score 600; DB 3; Length 151;
 Best Local Similarity 81.5%; Pred. No. 2.5e-50;
 Matches 123; Conservative 3; Mismatches 25; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
 DB 1 MKLLKVAAPAAIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
 QY 61 SDARKSETTITQSGYNGADVDQVTRVVTHEMAHAFRNATIDQNAKNSDITVGQYGG 120
 DB 61 SDARKSETTITQSGYNGADVDQVTRVVTHEMAHAFRNATIDQNAKNSDITVGQYGG 120
 QY 121 NNAALVNOTASDSSVWVRQVFGNNATANQY 151
 DB 121 HEMAHANOTASDSSVWVRQVFGNNATANQY 151

RESULT 13
 AAB36348

ID AAB36348 standard; protein; 151 AA.

XX AAB36348;

DT 26-FEB-2001 (first entry)

DE AGFA:PT3#3 amino acid sequence SEQ ID NO:16.

XX Salmonella; agfa; chromosomal gene replacement; fimbrial; epitope;

KW vaccine; immune response; immunogen.

XX Salmonella enteritidis.

OS *Escherichia coli*.

OS Synthetic.

XX WO200060102-A2.

XX 12-OCT-2000.

PF 05-APR-2000; 2000WO-CA000356.

PR 05-APR-1999; 99US-0127888P.

XX (UYVI-) UNIV VICTORIA.

PA White AP, Doran JL, Collison SK, Kay WW;

PI WPI; 2000-672631/65.

DR N-PSDB; AAC64624.

PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.

PS Disclosure; Page 136; 139pp; English.

CC The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAf) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

XX Sequence 151 AA;

Query Match 73.3%; Score 568; DB 3; Length 151;
 Best Local Similarity 80.1%; Pred. No. 3.3e-47;
 Matches 121; Conservative 4; Mismatches 26; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
 DB 1 MKLLKVAAPAAIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
 QY 61 SDARKSETTITQSGYNGADVDQVTRVVTHEMAHAFRNATIDQNAKNSDITVGQYGG 120
 DB 61 SDARKSETTITQSGYNGADVDQVTRVVTHEMAHAFRNATIDQNAKNSDITVGQYGG 120
 QY 121 NNAALVNOTASDSSVWVRQVFGNNATANQY 151
 DB 121 NNAALVNOTASDSSVWVRQVFGNNATANQY 151

RESULT 14
 AAB36343

ID AAB36343 standard; protein; 151 AA.

XX AAB36343;

DT 26-FEB-2001 (first entry)

DE *Escherichia coli* CsgA amino acid sequence SEQ ID NO:7.

KW *Salmonella*; agfa; chromosomal gene replacement; fimbrial; epitope;

KW vaccine; immune response; immunogen.

XX *Escherichia coli*.

XX WO200060102-A2.

PD 12-OCT-2000.

PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UTVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay MW;
XX WPI: 2000-672631/65.
DR N-PSDB; AAC64619.
XX
PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 135; 139pp; English.
XX
The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SFP17/TA) nucleation depended
CC assembly system of strains of *Salmonella*, *Escherichia coli* and
CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
CC *Enterobacteriaceae* host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 100,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 151 AA:
XX
Query Match 65.7%; Score 509; DB 3; Length 151;
Best Local Similarity 68.2%; Pred. No. 1.8e-41;
Matches 103; Conservative 15; Mismatches 33; Indels 0; Gaps 0;
QY 1 MKLLKVAAPAAIVVSGSALAGVPPQWGGGNNHNGGNSGSPDSTLSIYQGSANPAALAQ 60
DB 1 MKLLKVAALAAIVFSSSALAGVPPQYGGGNNHNGGNSGPNSEINITYQGGNSALALQ 60
QY 61 SDAKSETTITGSGYNGADYDQLVTRVVTHEMAHAFRNATTDQNNKXSDITVQYGG 120
DB 61 TDARNSDLTITTHGCGNGADVCGSDSSIDLTRGFGNSATLIDQNNKXSEMTVQFGG 120
QY 121 NNAALVNQTAASDSVWVROVFGNNATANQY 151
DB 121 GNGAAVVDQTAHNSVAVTVQVFGNNATAHQY 151
XX
RESULT 15
ID ABR82651 standard; protein; 151 AA.
XX
AC ABR82651;
XX
DT 04-DEC-2003 (first entry)
XX
DE E. coli CsgA subunit 15 kDa protein.
XX

KW Plasma protein; immune response; antibacterial; vaccine; gene therapy.
XX
OS *Escherichia coli*.
XX
PN WO2003064446-A2.
XX
XX 07-AUG-2003.
PD
PF 30-JAN-2003; 2003WO-EP000943.
XX
PR 31-JAN-2002; 2002GB-00002275.
XX
PA (HANS-) HANSA MEDICAL RES AB.
XX
PI Bioerck L, Olsen A, Wikstroem M, Herwald H;
XX WPI: 2003-646136/61.
DR N-PSDB; ACF36153.
XX
PT New isolated peptide capable of binding a mammalian plasma protein,
PT useful in the manufacture of a medicament for the prevention and/or
PT treatment of a bacterial infection, such as *Escherichia coli*, *Salmonella*
PT or *Shigella* infections.
XX
PS Disclosure; Page 41-42; 42pp; English.
XX
XX The invention relates to an isolated peptide capable of binding a
CC mammalian plasma protein or of generating an immune response in a mammal
CC selected from sequences shown in ABR82642, ABR82648-49. The peptide or
CC antibody is useful for treating a bacterial infection in a human or
CC animal or in the manufacture of a medicament for the prophylactic
CC treatment of a bacterial infection, such as *Escherichia coli*, *Salmonella*
CC or *Shigella* infection. The peptide that is immobilized on a solid support
CC is also useful as a reagent for determining the ability of a plasma
CC protein to bind to bacteria. The present sequence represents an *E. coli*
CC 15 kDa protein
XX
SQ Sequence 151 AA:
XX
Query Match 65.0%; Score 504; DB 7; Length 151;
Best Local Similarity 67.5%; Pred. No. 5.5e-41;
Matches 102; Conservative 15; Mismatches 34; Indels 0; Gaps 0;
QY 1 MKLLKVAAPAAIVVSGSALAGVPPQWGGGNNHNGGNSGSPDSTLSIYQGSANPAALAQ 60
DB 1 MKLLKVEALAAIVFSSSALAGVPPQYGGGNNHNGGNSGPNSEINITYQGGNSALALQ 60
QY 61 SDAKSETTITGSGYNGADYDQLVTRVVTHEMAHAFRNATTDQNNKXSDITVQYGG 120
DB 61 TDARNSDLTITTHGCGNGADVCGSDSSIDLTRGFGNSATLIDQNNKXSEMTVQFGG 120
QY 121 NNAALVNQTAASDSVWVROVFGNNATANQY 151
DB 121 GNGAAVVDQTAHNSVAVTVQVFGNNATAHQY 151
XX
Search completed: March 11, 2004, 18:33:46
Job time : 46.9 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:25:54 ; Search time 12.4 Seconds
(without alignments)
628,671 Million cell updates/sec

Title: US-09-543-407-28

Sequence: 1 MLLKVAAPAAIVSGSALA.....DSVMRVQVFGNNATANCY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/prodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/prodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/prodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/prodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/prodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/prodata/2/iaa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	678	87.5	151	1	US-08-233-788A-59
2	498	64.3	120	1	US-08-233-788A-57
3	85	11.0	975	4	US-09-328-352-4764
4	84.5	10.9	738	3	US-08-864-038A-3
5	84	10.8	943	4	US-09-056-556-204
6	84	10.8	943	4	US-09-072-596-199
7	84	10.8	943	4	US-09-477-135A-111
8	84	10.8	943	4	US-09-072-567-204
9	82	10.6	2123	3	US-08-968-685A-10
10	81.5	10.5	558	4	US-09-252-991A-30983
11	80	10.3	339	4	US-09-252-991A-32096
12	80	10.3	1207	4	US-09-489-039A-11518
13	79.5	10.3	745	4	US-09-336-115C-6
14	79.5	10.3	1415	4	US-09-252-991A-26438
15	78.5	10.1	714	4	US-09-841-786-4
16	78.5	10.1	3241	4	US-09-841-786-1
17	78	10.1	892	4	US-09-336-447A-5
18	78	10.1	918	4	US-09-300-650E-1
19	77.5	10.0	415	4	US-09-025-769B-280
20	77.5	10.0	518	3	US-09-043-123-2
21	77	9.9	1160	3	US-08-808-599A-24
22	76.5	9.9	873	4	US-09-336-447A-13
23	75.5	9.7	211	1	US-08-276-852-34
24	75.5	9.7	211	1	US-08-133-011-16
25	75.5	9.7	211	1	US-08-322-730A-16
26	75.5	9.7	211	1	US-08-387-874-16
27	75.5	9.7	211	1	US-08-899-575-34

28	75.5	9.7	211	1	US-08-899-575-34	Sequence 34, Appl
29	75.5	9.7	211	2	US-08-383-618-16	Sequence 16, Appl
30	75.5	9.7	211	3	US-08-907-733-16	Sequence 16, Appl
31	75.5	9.7	211	4	US-09-729-597-16	Sequence 16, Appl
32	75.5	9.7	211	5	PCT-US93-08364-16	Sequence 16, Appl
33	75.5	9.7	211	5	PCT-US95-08743-34	Sequence 34, Appl
34	75.5	9.7	238	4	US-09-495-880A-42	Sequence 42, Appl
35	75.5	9.7	244	2	US-08-553-497A-22	Sequence 22, Appl
36	75.5	9.7	266	4	US-09-495-880A-26	Sequence 26, Appl
37	75.5	9.7	293	3	US-08-438-745-4	Sequence 4, Appl
38	75.5	9.7	293	3	US-08-438-745-6	Sequence 6, Appl
39	75.5	9.7	293	3	US-09-219-019-4	Sequence 4, Appl
40	75.5	9.7	293	3	PCT-US94-05669A-4	Sequence 6, Appl
41	75.5	9.7	293	5	PCT-US94-05669A-6	Sequence 6, Appl
42	75.5	9.7	293	5	US-09-403-089A-1	Sequence 1, Appl
43	75.5	9.7	432	2	US-08-463-587A-26	Sequence 26, Appl
44	75.5	9.7	461	2	US-08-463-667A-4	Sequence 4, Appl
45	75.5	9.7	461	2	US-08-463-667A-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-08-233-788A-59
Sequence 59, Application US/08233788A

Patent No. 5635617

GENERAL INFORMATION:

APPLICANT: Doran, James L.

APPLICANT: Kay, William W.

APPLICANT: Collinson, Karen S.

APPLICANT: Clouthier, Sharon C.

TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION

NUMBER OF SEQUENCES: 61

CORRESPONDENCE ADDRESS:

ADDRESSER: Seed and Berry

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: U.S.A.

ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/233,788A

FILING DATE: 26-APR-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: King, Joshua

REGISTRATION NUMBER: 35,570

REFERENCE/DOCKET NUMBER: 920043.403C2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

TELEX: 3723836 SEDANBERY

INFORMATION FOR SEQ ID NO: 59:

SEQUENCE CHARACTERISTICS:

LENGTH: 151 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-233-788A-59

Query Match 87.5%; Score 678; DB 1; Length 151;

Best Local Similarity 89.4%; Pred. No. 46-61;

Matches 135; Conservative 1; Mismatches 15; Indels 0; Gaps 0;

QY 1 MLLKVAAPAAIVSGSALAGVFPWGSGGNGGNSGPPSTISITYGSANALAO 60

Db 1 MLLKVAFAALVVGSGALAGVPPWGGGNGGNSGSPDSTLSTIYOGSANAALALQ 60
 QY 61 SPARKSETTITGSGYGADYDQLVTRVVTHEMAHAFRNNAITIDMNAKNSDITVGYG 120
 Db 61 SPARKSETTITGSGYGADYDQLVTRVVTHEMAHAFRNNAITIDMNAKNSDITVGYG 120
 QY 121 NNAALVNCASSSVWVQVGFNNATANOY 151
 Db 121 NNPALVNOTASDSVWVQVGFNNATANOY 151

RESULT 2

US-08-233-788A-57
 ; Sequence 57, Application US/08233788A
 ; Patent No. 5635617
 ; GENERAL INFORMATION:
 ; APPLICANT: Doran, James L.
 ; APPLICANT: Kay, William W.
 ; APPLICANT: Collinson, Karen S.
 ; APPLICANT: Clouthier, Sharon C.
 ; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
 ; NUMBER OF SEQUENCES: 61
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Seed and Berry
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue
 ; CITY: Seattle
 ; STATE: Washington
 ; COUNTRY: U.S.A.
 ; ZIP: 98104-7092
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/233,788A
 ; FILING DATE: 26-APR-1994
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: King, Joshua
 ; REGISTRATION NUMBER: 35,570
 ; REFERENCE/DOCKET NUMBER: 920043.403C2
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (206) 622-4900
 ; TELEFAX: (206) 682-6031
 ; TELEX: 3723836 SEEDANBERRY
 ; INFORMATION FOR SEQ ID NO: 57:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 120 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-233-788A-57

Query Match 64.3%; Score 498; DB 1; Length 120;
 Best Local Similarity 86.6%; Pred. No. 4.6e-43;
 Matches 97; Conservative 1; Mismatches 14; Indels 0; Gaps 0;

QY 22 VPPWGGGNGGNSGSPDSTLSTIYOGSANAALALQSDARKSETTITGSGYGADY 81
 Db 1 VPPWGGGNGGNSGSPDSTLSTIYOGSANAALALQSDARKSETTITGSGYGADY 60
 QY 82 DQLVTRVVTHEMAHAFRNNAITIDMNAKNSDITVGYGNNALVNOTASDS 133
 Db 61 GGGADNSTIELTONGFRNNAITIDMNAKNSDITVGYGNNALVNOTASDS 112

RESULT 3
 US-09-328-352-4764
 ; Sequence 4764, Application US/09328352
 ; Patent No. 6562958
 ; GENERAL INFORMATION:

; APPLICANT: Gary L. Breton et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
 ; FILE REFERENCE: GTC99-03PA
 ; CURRENT APPLICATION NUMBER: US/09/328,352
 ; CURRENT FILING DATE: 1999-06-04
 ; NUMBER OF SEQ ID NOS: 8252
 ; SEQ ID NO 4764
 ; LENGTH: 975
 ; TYPE: PRT
 ; ORGANISM: Acinetobacter baumannii
 ; US-09-328-352-4764

Query Match 11.0%; Score 85; DB 4; Length 975;
 Best Local Similarity 25.3%; Pred. No. 4.2;
 Matches 38; Conservative 18; Mismatches 52; Indels 42; Gaps 7;

QY 15 GSGALAGVPPWGGGNGGNSGSPDSTLSTIYOG-----SANA 55
 Db 300 AGNGIA-----SGNGEHYIGNGGDDVDITAPITGLNISGNSFTLGNSSSSSVT 353
 QY 56 ALALQSDARKSETTI-----TOSGYGADYDQLVTRVVTHEMAHAFRNNAITIDMNA 108
 Db 354 APTTTSNTVNDTIDNGSGGTGSGSGSGG-DGLNGAASNGEHR---NYGIGNGG 408
 QY 109 KNSDIT-----VGYGNNALVNOTASDS 133
 Db 409 DVDITSPITGTFNFSGNSFSLIGNSSSS 438

RESULT 4

US-08-864-038A-3
 ; Sequence 3, Application US/08864038A
 ; Patent No. 6001592
 ; GENERAL INFORMATION:
 ; APPLICANT: Kunio NAKASHIMA et al
 ; TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE CDNA, VECTOR
 ; TITLE OF INVENTION: CONTAINING SAID CDNA, HOST CELLS TRANSFORMED WITH SAID
 ; TITLE OF INVENTION: VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING
 ; TITLE OF INVENTION: TO SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBODY
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: 812-5 Hirano
 ; STREET: Isshinden
 ; CITY: Tau-city
 ; STATE: Mie-prefecture
 ; COUNTRY: JAPAN
 ; ZIP: 514-01
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: Microsoft Windows 95
 ; SOFTWARE: Word Perfect 6.1
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/864,038A
 ; FILING DATE: May 28, 1997
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: JP 8-184459
 ; FILING DATE: 15-July-1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: C. Bruce Hamburg
 ; REGISTRATION NUMBER: 22,389
 ; REFERENCE/DOCKET NUMBER: F-5610
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212)986-2340
 ; TELEFAX: (212)953-7733
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 738
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein

ORIGINAL SOURCE:
ORGANISM: Pinctada fucata
CELL TYPE: mantle epithelial cell
FEATURE:
NAME/KEY: peptide
LOCATION: from 1 to 738
IDENTIFICATION METHOD: E (by experiment)
US-08-864-038A-3

Query Match 10.8%; Score 84.5; DB 3; Length 738;
Best Local Similarity 35.5%; Pred. No. 3.3;
Matches 27; Conservative 4; Mismatches 22; Indels 23; Gaps 3;

QY 3 LKVAAPAAIVGSGALAGVVPQMGCGGNHNGGNSGPGSTLSIYOGSANAALALQSD 62
DB 419 LKSSASASASASASASAG-----GGGGGNGGNGGG-----GGAGALA----- 460

QY 63 ARKSETTITQSGYNG 78
DB 461 -----NALAAGAGGG 471

RESULT 5

US-09-056-556-204
Sequence 204, Application US/09056556
Patent No. 6350456
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, David C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
NUMBER OF SEQUENCES: 241
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/056,556
FILING DATE: 07-Apr-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.457
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 204:
SEQUENCE CHARACTERISTICS:
LENGTH: 943 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-09-056-556-204

Query Match 10.8%; Score 84; DB 4; Length 943;
Best Local Similarity 24.7%; Pred. No. 5.1;
Matches 38; Conservative 12; Mismatches 54; Indels 50; Gaps 7;

QY 16 GSALAGVVPQMGCG-GNHN-GGNSGPGSTLSIYOGSANAALALQSDARKSETTITQS 73
DB 464 GSGNIGVFNVGSGSLGNYNGSGN-----LGNYNGFNGVG-----DYNV 503

QY 74 GYNGADYDQLVTRVVTHEMAHAFNNATIDQNNAKNSDITVGOYGN----- 121

DB 504 GFNGADFNQ-----GFANTGNNNIGFANTGNNNIGIGLSGDNQGFNIASGWS 553

QY 122 ---NALVNGTASDSSVM---VROVFGNATAN 149
DB 554 GTGNSGLFNSGTNNVGIIFNAGTGNVGIANSGTGN 587

RESULT 6

US-09-072-596-199
Sequence 199, Application US/09072596
Patent No. 6458366
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, David C.
APPLICANT: Campos-Neto, Antonia
APPLICANT: Houghton, Raymond
APPLICANT: Veddzik, Thomas S.
APPLICANT: Twardzik, Daniel R.
APPLICANT: Lodes, Michael J.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
NUMBER OF SEQUENCES: 350
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/072,596
FILING DATE: 05-MAY-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 199:
SEQUENCE CHARACTERISTICS:
LENGTH: 943 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-09-072-596-199

Query Match 10.8%; Score 84; DB 4; Length 943;
Best Local Similarity 24.7%; Pred. No. 5.1;
Matches 38; Conservative 12; Mismatches 54; Indels 50; Gaps 7;

QY 16 GSALAGVVPQMGCG-GNHN-GGNSGPGSTLSIYOGSANAALALQSDARSETTITQS 73
DB 464 GSGNIGVFNVGSGSLGNYNGSGN-----LGNYNGFNGVG-----DYNV 503

QY 74 GYNGADYDQLVTRVVTHEMAHAFNNATIDQNNAKNSDITVGOYGN----- 121
DB 504 GFNGADFNQ-----GFANTGNNNIGFANTGNNNIGIGLSGDNQGFNIASGWS 553

QY 122 ---NALVNGTASDSSVM---VROVFGNATAN 149
DB 554 GTGNSGLFNSGTNNVGIIFNAGTGNVGIANSGTGN 587

RESULT 7
US-09-477-135A-131

```

; Sequence 131, Application US/09477135A
; Patent No. 6572865
; GENERAL INFORMATION:
; APPLICANT: NANO, Francis
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
; FILE REFERENCE: 52888
; CURRENT APPLICATION NUMBER: US/09/477,135A
; CURRENT FILING DATE: 2000-01-03
; PRIOR APPLICATION NUMBER: 08990823
; PRIOR FILING DATE: 1997-12-15
; PRIOR APPLICATION NUMBER: US 96/10375
; PRIOR FILING DATE: 1996-06-14
; PRIOR APPLICATION NUMBER: 60/000,254
; NUMBER OF SEQ ID NOS: 169
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 131
; LENGTH: 943
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-477-135A-131

```

```

Query Match      10.8%; Score 84; DB 4; Length 943;
Best Local Similarity 24.7%; Pred. No. 5.1;
Matches 38; Conservative 12; Mismatches 54; Indels 50; Gaps 7;

QY 16 GSALAGVFPWQGG-GNHN-GGNSGSPDSTLSIYQGSANAALALQSDARKSETTITQS 73
DB 571 GSGNIGVFNVGSGSLGNINIGSGN-----LGIYIGFNVG-----DNYV 610
QY 74 GYNGADYDQLVTRVTHEMAHAFRNATIDQWNAKSDITVQYQGN-----121
DB 611 GFNAGDFNQ-----GFANTGNNNIGFANTGNNNIGIGLSGDNQGFNIASGWS 660
QY 122 ---NALVNOTASDSSVW---VROYGFNATATN 149
DB 661 GTNSGLFNSGTNNVGIFNAGTGNVGIANSGTGN 694

```

```

RESULT 8
US-09-072-967-204
; Sequence 204, Application US/09072967
; Patent No. 6592877
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:

```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 204:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 943 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-09-072-967-204

```

```

Query Match      10.8%; Score 84; DB 4; Length 943;
Best Local Similarity 24.7%; Pred. No. 5.1;
Matches 38; Conservative 12; Mismatches 54; Indels 50; Gaps 7;

QY 16 GSALAGVFPWQGG-GNHN-GGNSGSPDSTLSIYQGSANAALALQSDARKSETTITQS 73
DB 464 GSGNIGVFNVGSGSLGNINIGSGN-----LGIYIGFNVG-----DNYV 503
QY 74 GYNGADYDQLVTRVTHEMAHAFRNATIDQWNAKSDITVQYQGN-----121
DB 504 GFNAGDFNQ-----GFANTGNNNIGFANTGNNNIGIGLSGDNQGFNIASGWS 553
QY 122 ---NALVNOTASDSSVW---VROYGFNATATN 149
DB 554 GTNSGLFNSGTNNVGIFNAGTGNVGIANSGTGN 587

```

```

RESULT 9
US-08-968-685A-10
; Sequence 10, Application US/08968685A
; Patent No. 6214981
; GENERAL INFORMATION:
; APPLICANT: TUCKER, KENNETH
; APPLICANT: PLOSILA, LAURA
; TITLE OF INVENTION: MORAXELLA CATARRHALIS OUTER MEMBRANE
; TITLE OF INVENTION: PROTEIN-106 POLYPEPTIDE, GENE SEQUENCE AND USES THEREOF
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/968,685A
; FILING DATE: No. 6214981ember 12, 1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Baldwin, Geraldine F.
; REGISTRATION NUMBER: 31,232
; REFERENCE/DOCKET NUMBER: 7969-060
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2123 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide

```



```

; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)...(20)
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 721
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-336-115C-6

```

```

Query Match      10.3%; Score 79.5; DB 4; Length 745;
Best Local Similarity 20.5%; Pred. No. 11;
Matches 35; Conservative 27; Mismatches 70; Indels 39; Gaps 6;

```

```

QY 8 AFAIVVSGSALAGVPPQW-----GGGNNHGGGN----- 37
DB 102 AYQAVFLINAVGL---WNTIGYAVMCGNNGNGSTESGPGSYIFPDQRODSTQITCNRF 158
QY 38 SSGPDSTLSIYQGSANALALQSDARKSETTITQSGYGADYDQLVTRVTHMAHAF 97
DB 159 STGFGKMSIDFEPKLINEAYQIIQALKNQSGPELG-GNGTK-----VSVMNYECRQTA 213
QY 98 RNNATIDQW-NAKNSDITVQYGGNNALVNOTASDSSVMVRQVGFNNAT 147
DB 214 DINGGVYFCRAKNGSSSSNGNGSSTQTTATTQDGVITTTTNNKAT 264

```

```

RESULT 14
US-09-252-991A-26438
; Sequence 26438, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 26438
; LENGTH: 1415
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-26438

```

```

Query Match      10.3%; Score 79.5; DB 4; Length 1415;
Best Local Similarity 28.9%; Pred. No. 25;
Matches 44; Conservative 17; Mismatches 74; Indels 17; Gaps 9;

```

```

QY 10 AAIYVSGSALAGVPPQWGGG--NHNGG--NSGPDSTLSIYQGSANALALQSDA-- 63
DB 930 ADFALISQ-LKDHASHYGAGGLVGRNNGGLTRSSGSGQTLTSLSGHNNLGGLYGSSAGG 988
QY 64 -RKSETTITQSGYGADYDQLVTRVTHMAHAFRNNATIDQWNAKNSDITVQYGGN- 121
DB 989 LADVASAVDYSGNGQRLYGLIGLIVNSGIAHAHTASGKV---RGTDAELGGLIGRNL 1044
QY 122 NAALVNOTA--SDSSVMV-RQVG--FGNNATAN 149
DB 1045 NAAINNASAHGVSLOAGRYLGLIGHNQAQN 1076

```

```

RESULT 15
US-09-841-786-4
; Sequence 4, Application US/09841786
; Patent No. 6669940
; GENERAL INFORMATION:
; APPLICANT: NAGARAJA, T. G.
; APPLICANT: STEWART, GEORGE C.
; APPLICANT: NARAYANAN, SANJEEV K.

```

```

; APPLICANT: CHENGAPPA, M. M.
; TITLE OF INVENTION: RECOMBINANT FUSOBACTERIUM NECROPHORUM LEUKOTOXIN
; TITLE OF INVENTION: VACCINE AND PREPARATION THEREOF
; FILE REFERENCE: 30296
; CURRENT APPLICATION NUMBER: US/09/841,786
; CURRENT FILING DATE: 2001-04-24
; PRIOR APPLICATION NUMBER: 09/558,257
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 714
; TYPE: PRT
; ORGANISM: Fusobacterium necrophorum
US-09-841-786-4

```

```

Query Match      10.1%; Score 78.5; DB 4; Length 714;
Best Local Similarity 22.8%; Pred. No. 13;
Matches 31; Conservative 16; Mismatches 56; Indels 33; Gaps 6;

```

```

QY 12 IYVSGSALAGVPPQWGGG-----NHNGGN-----SSGPDSTLSIYQGSANALALQSDAR 64
DB 35 VTSSDSTFVGA---WGSALQWNNHIGSGNSNISAGLAAAVNNIQSTSLVKNSDIR 91
QY 65 KSE-----TTITQSGYGADY---DQLVTRVTHMAHAFNNAT 102
DB 92 NANKFKVNLSSGTQVAAGAGLEAVKESG-GQKSYLTGTSAINLVNNEVSAKSENNTV 150
QY 103 IDQWNAKNSDITVQY 118
DB 151 AGESESQKMDVDTAY 166

```

```

Search completed: March 11, 2004, 18:44:54
Job time : 13.4 secs

```

Fri Mar 12 09:42:03 2004

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:40:56 ; Search time 24.6 Seconds
(without alignments)
1296.103 Million cell updates/sec

Title: US-09-543-407-28

Perfect score: 775
Sequence: 1 MGLKVAAPALVSGSALA.....DSVWVROYFGNNATANDY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 809742 seqs, 21153259 residues

Total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/FCR_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	89	11.5	354	10	US-09-820-843A-21
2	87.5	11.3	249	9	US-09-730-374-3
3	87.5	11.3	445	15	US-10-369-493-20638
4	87	11.2	597	9	US-09-793-306-146
5	85	11.0	254	10	US-09-880-748-905
6	84	10.8	943	9	US-09-996-634-131
7	84	10.8	943	10	US-09-997-182-111
8	84	10.8	943	10	US-09-997-181-131
9	84	10.8	943	14	US-10-193-002-189
10	84	10.8	943	14	US-10-084-843-204
11	82	10.6	2122	9	US-09-813-214A-9
12	81	10.5	224	9	US-09-738-626-4894
13	81	10.5	812	15	US-10-369-493-9134
14	80	10.3	246	10	US-09-880-748-1647
15	80	10.3	252	10	US-09-880-748-1201

15	80	10.3	253	10	US-09-880-748-2098	Sequence 2098, Ap
17	79.5	10.3	278	9	US-09-810-264-28	Sequence 28, Appl
18	79.5	10.3	745	8	US-08-834-666A-6	Sequence 6, Appl1
19	78.5	10.2	65	9	US-09-996-194-16	Sequence 16, Appl
20	78.5	10.1	486	15	US-10-369-493-20619	Sequence 20619, A
21	78.5	10.1	714	9	US-09-841-786-4	Sequence 4, Appl1
22	78.5	10.1	3241	9	US-09-841-786-1	Sequence 1, Appl1
23	78	10.1	892	10	US-09-952-267-5	Sequence 5, Appl1
24	77.5	10.0	438	14	US-10-156-761-9343	Sequence 9343, Ap
25	77.5	10.0	518	9	US-09-976-297-2	Sequence 2, Appl1
26	77.5	10.0	562	14	US-10-156-761-13039	Sequence 13039, A
27	77.5	10.0	1621	14	US-10-185-990-10	Sequence 10, Appl
28	77.5	10.0	1626	14	US-10-185-990-11	Sequence 11, Appl
29	77	9.9	250	14	US-10-072-301-21	Sequence 21, Appl
30	77	9.9	250	14	US-10-072-301-29	Sequence 29, Appl
31	77	9.9	250	14	US-10-071-866-21	Sequence 21, Appl
32	77	9.9	250	14	US-10-071-866-29	Sequence 29, Appl
33	77	9.9	250	15	US-10-360-828-21	Sequence 21, Appl
34	77	9.9	250	15	US-10-360-828-29	Sequence 29, Appl
35	77	9.9	255	10	US-09-880-748-1153	Sequence 1153, Ap
36	77	9.9	255	10	US-09-880-748-1862	Sequence 1862, Ap
37	77	9.9	257	10	US-09-880-748-1494	Sequence 1494, Ap
38	77	9.9	271	14	US-10-156-761-11721	Sequence 11721, A
39	77	9.9	1046	15	US-10-369-493-1547	Sequence 1547, Ap
40	76.5	9.9	251	10	US-09-880-748-711	Sequence 71, Appl
41	76.5	9.9	608	14	US-10-156-761-13874	Sequence 13874, A
42	76.5	9.9	873	10	US-09-952-267-13	Sequence 13, Appl
43	76.5	9.9	1088	15	US-10-421-654-72	Sequence 72, Appl
44	76.5	9.9	1778	14	US-10-238-075-749	Sequence 749, Ap
45	76.5	9.9	2732	14	US-10-238-075-1119	Sequence 1119, Ap

ALIGNMENTS

RESULT 1
US-09-820-843A-21
; Sequence 21, Application US/09820843A
; Publication No. US20030039963A1
; GENERAL INFORMATION:
; APPLICANT: Council of Scientific and Industrial Research
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEI
; FILE REFERENCE: 063915
; CURRENT APPLICATION NUMBER: US/09/820,843A
; CURRENT FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 21
; LENGTH: 354
; TYPE: PRT
; ORGANISM: M. tuberculosis
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: PPS
; NAME/KEY: misc feature
; OTHER INFORMATION: 911781260
US-09-820-843A-21

Query Match 11.5%; Score 89; DB 10; Length 354;
Best Local Similarity 24.1%; Pred. No. 1;
Matches 38; Conservative 17; Mismatches 63; Indels 40; Gaps 7;
QY 10 AAIIVSGSALG-VYPQWGGGNNHGGG-----NSSGPPSTLSIYOYGSAN 54
DB 200 AGDVIVSGVGNAGDVVTGIGNSNITGTFNPGTNTGTFSSAMTOGCPNS--CFNAGTGN 257
QY 55 PALALQSDARKSETTITSGYGNAGDYDQLVYRVVTHEMAHAFRNNAITDQNAKNSDIT 114
DB 258 SGFGHNDPAVGSNGSIGNSGFGNS-----GYVITSTSMGSGSVLGN 300
QY 115 VGQYGNNAALVNOTASDVSVWVROY---GFGNNATAN 149

PRIOR FILING DATE: 2000-06-15
PRIOR APPLICATION NUMBER: 60/240,816
PRIOR FILING DATE: 2000-10-17
PRIOR APPLICATION NUMBER: 60/276,248
PRIOR FILING DATE: 2001-03-16
PRIOR APPLICATION NUMBER: 60/277,379
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/293,499
PRIOR FILING DATE: 2001-05-25
NUMBER OF SEQ ID NOS: 3239
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 905
LENGTH: 254
TYPE: PRT
ORGANISM: Homo sapiens
US-09-880-748-905

Query Match 11.0%; Score 85; DB 10; Length 254;
Best Local Similarity 23.0%; Pred. No. 1.7;
Matches 29; Conservative 11; Mismatches 42; Indels 44; Gaps 3;

QY 13 VVSGSLAGVPPQMG-----GGNNHNGGNSGPDSTLSI-----47
DB 104 ILTGYYMGAFFDQMGRTWTVSSGGGSGGSGGSLLETVTQSPCTLSLSPGERA 163
QY 48 -----YQYGSANALALQSDARKSETTTQ--SGYNGADYDQLVTRY 88
DB 164 TUSCRASQSIIRSNYLAWYQKSGQAPRLIYDVSRBATGIPDRFSGSGSGTDFLTISRL 223
QY 89 VTHEMA 94
DB 224 EPEDFA 229

RESULT 6
US-09-996-634-131
Sequence 131, Application US/09996634
Patent No. US20020172684A1
GENERAL INFORMATION:
APPLICANT: Nano, Francis
TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
TITLE OF INVENTION: Immunostimulatory Peptides
FILE REFERENCE: 61260
CURRENT APPLICATION NUMBER: US/09/996,634
CURRENT FILING DATE: 2001-11-28
PRIOR APPLICATION NUMBER: 09/447,135
PRIOR FILING DATE: 2000-01-03
PRIOR APPLICATION NUMBER: 08/990,823
PRIOR FILING DATE: 1997-12-15
PRIOR APPLICATION NUMBER: US 96/10375
PRIOR FILING DATE: 1996-06-14
PRIOR APPLICATION NUMBER: 60/000,254
PRIOR FILING DATE: 1995-06-15
NUMBER OF SEQ ID NOS: 169
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 131
LENGTH: 943
TYPE: PRT
ORGANISM: Mycobacterium tuberculosis
US-09-996-634-131

Query Match 10.8%; Score 84; DB 9; Length 943;
Best Local Similarity 24.7%; Pred. No. 11;
Matches 38; Conservative 12; Mismatches 54; Indels 50; Gaps 7;
QY 16 GSAALAGVPPQMGG-GNNH-GGNSGPDSTLSIYQYGSANALALQSDARKSETTTQ 73
DB 571 GSGNIGVFVNGSGSLGNVNGSGN-----LGIYNGGNGV-----DYNV 610
QY 74 GYNGADYDQLVTRYVTHEMAHAFRNATIDQNNAKNSDITVGOYGN-----121
DB 611 GFNGAGDFNQ-----GFANTGNNNIGFANTGNNNIGIGLSDNOQGFIASGMS 660

QY 122 ---NALVNGTASDSSVM---VROYFGNNATAN 149
DB 661 GTNGSGLFNSGTNNVGFNAGTGNVGIANSGTGN 694

RESULT 7
US-09-997-182-131
Sequence 131, Application US/09997182
Publication No. US20030049263A1
GENERAL INFORMATION:
APPLICANT: Nano, Francis
TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
TITLE OF INVENTION: Immunostimulatory Peptides
FILE REFERENCE: 61258
CURRENT APPLICATION NUMBER: US/09/997,182
CURRENT FILING DATE: 2001-11-28
PRIOR APPLICATION NUMBER: 09/447,135
PRIOR FILING DATE: 2000-01-03
PRIOR APPLICATION NUMBER: 08/990,823
PRIOR FILING DATE: 1997-12-15
PRIOR APPLICATION NUMBER: US 96/10375
PRIOR FILING DATE: 1996-06-14
PRIOR APPLICATION NUMBER: 60/000,254
PRIOR FILING DATE: 1995-06-15
NUMBER OF SEQ ID NOS: 169
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 131
LENGTH: 943
TYPE: PRT
ORGANISM: Mycobacterium tuberculosis
US-09-997-182-131

Query Match 10.8%; Score 84; DB 10; Length 943;
Best Local Similarity 24.7%; Pred. No. 11;
Matches 38; Conservative 12; Mismatches 54; Indels 50; Gaps 7;

QY 16 GSAALAGVPPQMGG-GNNH-GGNSGPDSTLSIYQYGSANALALQSDARKSETTTQ 73
DB 571 GSGNIGVFVNGSGSLGNVNGSGN-----LGIYNGGNGV-----DYNV 610
QY 74 GYNGADYDQLVTRYVTHEMAHAFRNATIDQNNAKNSDITVGOYGN-----121
DB 611 GFNGAGDFNQ-----GFANTGNNNIGFANTGNNNIGIGLSDNOQGFIASGMS 660
QY 122 ---NALVNGTASDSSVM---VROYFGNNATAN 149
DB 661 GTNGSGLFNSGTNNVGFNAGTGNVGIANSGTGN 694

RESULT 8
US-09-997-181-131
Sequence 131, Application US/09997181
Publication No. US20030049269A1
GENERAL INFORMATION:
APPLICANT: Nano, Francis
TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
TITLE OF INVENTION: Immunostimulatory Peptides
FILE REFERENCE: 61257
CURRENT APPLICATION NUMBER: US/09/997,181
CURRENT FILING DATE: 2001-11-28
PRIOR APPLICATION NUMBER: 09/447,135
PRIOR FILING DATE: 2000-01-03
PRIOR APPLICATION NUMBER: 08/990,823
PRIOR FILING DATE: 1997-12-15
PRIOR APPLICATION NUMBER: US 96/10375
PRIOR FILING DATE: 1996-06-14
PRIOR APPLICATION NUMBER: 60/000,254
PRIOR FILING DATE: 1995-06-15
NUMBER OF SEQ ID NOS: 169
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 131
LENGTH: 943
TYPE: PRT

ORGANISM: Mycobacterium tuberculosis
US-09-997-181-131

Query Match 10.8%; Score 84; DB 10; Length 943;
Best Local Similarity 24.7%; Pred. No. 11;
Matches 38; Conservative 12; Mismatches 54; Indels 50; Gaps 7;

QY 16 GSALAGVVPQWGG-GNNH-GGNSSGPDSLTIYQGSANALALQSDARKSETTITQS 73
DB 571 GSGNIGCVFNVGSGSLGNVNISSGN-----LGIYNIQGVNG-----DYNV 610
QY 74 GYGNAGDYDQLVTRVYTHEMAHAFRNNAITIDQNAKNSDITVQYGN-----121
DB 611 GFNAGDFNQ-----GFANTGNNNIGFANTGNNNIGIGLSGDNQGFNIASGWS 660
QY 122 ---NAALVQNTASDSSVM---VRQVFGNNAATAN 149
DB 661 GTGNSGLFNSTGNVNIQFNACTGNVGIANSCTGN 694

RESULT 9

US-10-193-002-199
Sequence 199, Application US/10193002
Publication No. US20030135026A1

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
Skeiky, Yasir A.W.
Dillon, Davin C.
Campos-Neto, Antonio
Houghton, Raymond
Vedvick, Thomas S.
Twardzik, Daniel R.
Lodes, Michael J.
Hendrickson, Ronald C.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
NUMBER OF SEQUENCES: 350
CORRESPONDENCE ADDRESS:
ADDRESS: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/193,002
FILING DATE: 10-Jul-2002
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/072,596
FILING DATE: 05-MAY-1998

ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392

TELECOMMUNICATION INFORMATION:
REFERENCE/DOCKET NUMBER: 210121.417C9
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 199:
SEQUENCE CHARACTERISTICS:
LENGTH: 943 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>

TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 199:
US-10-193-002-199

Query Match 10.8%; Score 84; DB 14; Length 943;

Best Local Similarity 24.7%; Pred. No. 11;
Matches 38; Conservative 12; Mismatches 54; Indels 50; Gaps 7;

QY 16 GSALAGVVPQWGG-GNNH-GGNSSGPDSLTIYQGSANALALQSDARKSETTITQS 73
DB 464 GSGNIGCVFNVGSGSLGNVNISSGN-----LGIYNIQGVNG-----DYNV 503
QY 74 GYGNAGDYDQLVTRVYTHEMAHAFRNNAITIDQNAKNSDITVQYGN-----121
DB 504 GFNAGDFNQ-----GFANTGNNNIGFANTGNNNIGIGLSGDNQGFNIASGWS 553
QY 122 ---NAALVQNTASDSSVM---VRQVFGNNAATAN 149
DB 554 GTGNSGLFNSTGNVNIQFNACTGNVGIANSCTGN 587

RESULT 10

US-10-084-843-204
Sequence 204, Application US/10084843
Publication No. US20030143243A1

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
Skeiky, Yasir A.W.
Dillon, Davin C.
Campos-Neto, Antonio
Houghton, Raymond
Vedvick, Thomas S.
Twardzik, Daniel R.
Lodes, Michael J.
Hendrickson, Ronald C.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 355
CORRESPONDENCE ADDRESS:
ADDRESS: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/084,843
FILING DATE: 25-Feb-2002
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/072,967
FILING DATE: 05-MAY-1998

ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C9

TELECOMMUNICATION INFORMATION:
REFERENCE/DOCKET NUMBER: 210121.411C9
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 204:
SEQUENCE CHARACTERISTICS:
LENGTH: 943 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>

TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 204:
US-10-084-843-204

Query Match 10.8%; Score 84; DB 14; Length 943;
Best Local Similarity 24.7%; Pred. No. 11;
Matches 38; Conservative 12; Mismatches 54; Indels 50; Gaps 7;

QY 16 GSALAGVVPQWGG-GNNH-GGNSSGPDSLTIYQGSANALALQSDARKSETTITQS 73

```

Db      464 GSGNIGVFNSSGSLGNINISGN-----LGITNIGFVG-----DYNV 503
Qy      74 GYGAGADYDQVTRVVTHEMAHAFRNATIDQMAKNSDITVGYGNN-----121
Db      504 GFGNAGDFNQ-----GFANTGNNNIGFANTGNNNIGISLSDNQGFNIASGWN 553
Qy      122 ---NAALVNQTAASDSSVW---VROYGFQNNATN 149
Db      554 GTGNSGLFNSTGNVGIPTNAGTGNVGIANSSTGN 587

```

RESULT 11
US-09-813-214A-9

```

; Sequence 9, Application US/09813214A
; Patent No. US20020177200A1
; GENERAL INFORMATION:
; APPLICANT: Tucker, Kenneth
; APPLICANT: Plesch, Laura
; TITLE OF INVENTION: MORAXELLA CATARRHALIS OUTER MEMBRANE PROTEIN-106 POLYPEPTIDE, GEN
; FILE REFERENCE: 7969-089-999
; CURRENT FILING DATE: 2000-03-20
; PRIOR APPLICATION NUMBER: US/09/813,214A
; PRIOR FILING DATE: 1997-11-12
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 9
; LENGTH: 2122
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-813-214A-9

```

Query Match 10.6%; Score 82; DB 9; Length 2122;
Best Local Similarity 25.8%; Pred. No. 48;
Matches 31; Conservative 17; Mismatches 38; Indels 34; Gaps 5;

```

Qy      1 MLLKVAFAFAIV-----SSSALAGVVPQWGGGNNHNGGSSGPDSTLSIYOGSANA 56
Db      44 LSFARIALAVLVIGATLNGSAVAGIGISEADGG--KGANARG--GDLAQ 97
Qy      57 LAQSDARKSETTTTQSGYGAGADYDQVTRVVTHEMAHAFRNATIDQMAKNSDITV 116
Db      98 LGSQSI-----IGDKIVHNSNNANIGAKASGESIAIG 133

```

RESULT 12
US-09-738-626-4894

```

; Sequence 4894, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059

```

```

; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 4894
; LENGTH: 224
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-4894

```

Query Match 10.5%; Score 81; DB 9; Length 224;
Best Local Similarity 24.6%; Pred. No. 3.7;
Matches 35; Conservative 19; Mismatches 54; Indels 34; Gaps 4;

```

Qy      16 GSALAGVVPQWGGGNNHNGGNS---GPDSTLSIYOGSANAALQDARSETTIT 71
Db      45 GTAIAGVLTGKAGATVEHIGSADSDIATNGDVVLAVPYRIVESIASHKDALAGKVID 104
Qy      72 QSGYGAGADYDQV-----TRVVTHEMAHAFRN--ATI-----103
Db      105 ITNPLNFETDLSLVPVGSATAEIQQLPTSRVLKFPNTNPAATLATGKVDITTVLV 164
Qy      104 --DQMAKNSDITVGYGNN 123
Db      165 AGDDEDAKNALITDVNAGLDA 186

```

RESULT 13
US-10-369-493-9134

```

; Sequence 9134, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 9134
; LENGTH: 812
; TYPE: PRT
; ORGANISM: Chloroflexus aurantiacus
US-10-369-493-9134

```

Query Match 10.5%; Score 81; DB 15; Length 812;
Best Local Similarity 26.6%; Pred. No. 18;
Matches 29; Conservative 9; Mismatches 31; Indels 40; Gaps 5;

```

Qy      14 VSGSALAGVVPQ-----WGGGNNHNGGNSGPDSTLSIYOGSANA 55
Db      696 TGSRSAGFSPDRVRSVPDAIGKVLDPYKNGNGNGNGNGHD---VEYNGE- 749
Qy      56 ALAQSDARKSETTTTQSGYGAGADYDQVTRVVTHEMAHAFRNATID 104
Db      750 -----TDTSRVTTNGAS--BATATTVSH--PHSDHGHGITE 782

```

RESULT 14
US-09-880-748-1847

```

; Sequence 1847, Application US/09880748
; Publication No. US20030059937A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Antibodies that Immunosepecifically Bind Blys
; FILE REFERENCE: PF523
; CURRENT APPLICATION NUMBER: US/09/880,748
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 60/212,210
; PRIOR FILING DATE: 2000-06-15

```

Search completed: March 11, 2004, 19:18:40
Job time : 25.6 secs

;; PRIOR APPLICATION NUMBER: 60/240,816
;; PRIOR FILING DATE: 2000-10-17
;; PRIOR APPLICATION NUMBER: 60/276,248
;; PRIOR FILING DATE: 2001-03-16
;; PRIOR APPLICATION NUMBER: 60/277,379
;; PRIOR FILING DATE: 2001-03-21
;; PRIOR APPLICATION NUMBER: 60/293,499
;; PRIOR FILING DATE: 2001-05-25
;; NUMBER OF SEQ ID NOS: 3239
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 1847
;; LENGTH: 246
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-09-880-748-1847

Query Match 10.3%; Score 80; DB 10; Length 246;
Best Local Similarity 31.1%; Pred. No. 5.2;
Matches 28; Conservative 6; Mismatches 32; Indels 24; Gaps 4;

QY 7 AAFPAIVSGSALAGVVPWG-----GGGNHGGNSGPDSTLSIYGSANAL 57
DB 88 AADTAVFCAATGKGYNDMNGRGMVTVSSGGSGGSGGSGG-----GSAQAVL 137
QY 58 ALQSD--ARKSETTITQSGYGN--GADYD 82
DB 138 TOPSSVSGAPGQRTTISCTGSSSNIGADYD 167

RESULT 15

US-09-880-748-1201
;; Sequence 1201, Application US/09880748
;; Publication No. US200305937A1
;; GENERAL INFORMATION:
;; APPLICANT: Ruben et al.
;; TITLE OF INVENTION: Antibodies that Immunospecifically Bind Blys
;; FILE REFERENCE: P5523
;; CURRENT APPLICATION NUMBER: US/09/880,748
;; CURRENT FILING DATE: 2001-06-15
;; PRIOR APPLICATION NUMBER: 60/212,210
;; PRIOR FILING DATE: 2000-06-15
;; PRIOR APPLICATION NUMBER: 60/240,816
;; PRIOR FILING DATE: 2000-10-17
;; PRIOR APPLICATION NUMBER: 60/276,248
;; PRIOR FILING DATE: 2001-03-16
;; PRIOR APPLICATION NUMBER: 60/277,379
;; PRIOR FILING DATE: 2001-03-21
;; PRIOR APPLICATION NUMBER: 60/293,499
;; PRIOR FILING DATE: 2001-05-25
;; NUMBER OF SEQ ID NOS: 3239
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 1201
;; LENGTH: 252
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-09-880-748-1201

Query Match 10.3%; Score 80; DB 10; Length 252;
Best Local Similarity 27.3%; Pred. No. 5.4;
Matches 42; Conservative 12; Mismatches 50; Indels 50; Gaps 9;

QY 26 WG-----GGGNHGGNSGPDSTLSIYGSANALALQSD--ARKSETTITQS 73
DB 114 WKGITLVTVSSGGSGGSGGSGG-----GSAQAVLTOPSSVSGAPGQRTTISCT 163
QY 74 GYGN--GADYDQVTRVVTHEMAH-----AFRNN-----ATIDOWNAKNSDIT--- 114
DB 164 GSGSNIGAGYD-----VHWYQHLPGTAPKLIIFNNNRPSGVPRFSGSKGTSASLA 216
QY 115 -VGQYGNNAALVNOTASDSVVRQVFGNNAT 147
DB 217 ITGLQGBDEADYVYQ--SFDTSLGVRVGGGTQILT 249

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:24:14 ; Search time 10.3 Seconds

(without alignments)
1410.186 Million cell updates/sec

Title: US-09-543-407-28

Perfect score: 775

Sequence: 1 MKLLKVAFAAIVSGSALA.....DSSVMVRQVGFNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	683	88.1	151	2 JC6039	fimbria protein ag
2	683	88.1	151	2 A10635	major curlin chain
3	509	65.7	151	2 S70788	curlin protein csg
4	487.5	62.9	152	2 D90806	curlin major subun
5	487.5	62.9	152	2 H85665	hypothetical prote
6	107	13.8	2174	2 E95965	hypothetical glyci
7	97	12.5	1748	2 S42136	cnjB protein - Tec
8	94.5	12.2	151	2 JC6040	fimbria protein ag
9	94	12.1	645	2 F70825	probable PPE prote
10	91.5	11.8	602	1 PL0221	leishmanolysin (BC
11	91	11.7	573	2 C86266	F319.21 protein -
12	89.5	11.5	151	2 S70787	probable PPE prote
13	89.5	11.5	151	2 C90806	curlin nucleator p
14	89.5	11.5	151	2 C90806	minor curlin subun
15	89.5	11.5	151	2 G85665	curlin minor chain
16	89.5	11.5	1322	2 S07053	ice nucleation pro
17	89.5	11.5	1567	2 S11672	ice nucleation pro
18	89	11.5	354	2 B70663	probable PPE prote
19	88	11.4	407	2 T21956	hypothetical prote
20	87.5	11.3	447	2 G84687	probable disease r
21	87.5	11.3	1034	2 JC2143	ice nucleation act
22	87.5	11.3	1258	2 JQ0188	ice nucleation pro
23	87	11.2	590	2 E70946	probable PPE prote
24	86.5	11.2	145	2 AD3143	conserved hypochet
25	86.5	11.2	145	2 H98144	hypothetical prote
26	86.5	11.2	151	2 AH0635	nucleation compone
27	86	11.1	590	1 A45621	leishmanolysin (EC
28	85.5	11.0	1053	2 B70987	probable PPE prote
29	85.5	11.0	2204	2 A70524	probable PPE prote

30	85	11.0	1052	2 AF2959	conserved hypochet
31	85	11.0	1341	2 H98323	hypothetical prote
32	84.5	10.9	646	1 S19916	leishmanolysin (EC
33	84.5	10.9	928	2 C81265	probable lipoprote
34	84.5	10.9	978	2 D81411	probable lipoprote
35	84	10.8	586	2 T26667	hypothetical prote
36	84	10.8	599	2 B42049	leishmanolysin (EC
37	84	10.8	599	2 A44951	leishmanolysin (EC
38	84	10.8	1608	2 A28182	hemolysin A - Ser
39	84	10.8	3300	2 D70575	probable PPE prote
40	83.5	10.8	582	2 F70675	probable PPE prote
41	83	10.7	277	2 AB1390	hypothetical cell
42	83	10.7	343	2 T05221	hypothetical prote
43	82.5	10.6	770	2 T51024	related to C2H2 z1
44	82.5	10.6	940	2 D89723	protein F39D8.1b l
45	82.5	10.6	945	2 T21998	hypothetical prote

ALIGNMENTS

RESULT 1
JC6039
fimbria protein agfa precursor - Salmonella enteritidis
C/Species: Salmonella enteritidis
C/Date: 31-Dec-1996 #sequence_revision 31-Dec-1996 #ext_change 08-Oct-1999
C/Accession: JC6039; PC6015; A44898
J/Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Baner, P.A.; Kay, W.W.
J. Bacteriol. 178, 662-667, 1996
A/Title: Salmonella enteritidis agfAFC operon encoding thin, aggregative fimbriae.
A/Reference number: JC6039; PMID:96146512; PMID:8550497
A/Accession: JC6039
A/Molecule type: DNA
A/Residues: 1-151 <COL>
A/Cross-references: GB:U43280; NID:G1184712; PIDN:ACG43599.1; PID:G1184714
A/Accession: PC6015
A/Molecule type: protein
A/Residues: 21-52 <CO2>
A/Experimental source: strain 27655-3b
A/Note: the authors translated the codon ACG for residue 44 as Ile
R/Collinson, S.K.; Emdy, L.; Muller, K.H.; Trust, T.J.; Kay, W.W.
J. Bacteriol. 173, 4773-4781, 1991
A/Title: Purification and characterization of thin, aggregative fimbriae from Salmonella
A/Reference number: A44898; PMID:91310586; PMID:1677357
A/Contents: 27655
A/Accession: A44898
A/Status: preliminary
A/Molecule type: protein
A/Residues: 21-33 <CO3>
A/Note: sequence extracted from NCBI backbone (NCBIP:45936)
C/Genetics:
A/Gene: agfa
C/Function:
A/Description: major component of thin aggregative fimbriae
A/Note: fimbriae bind to fibronectin, plasminogen, tissue plasminogen activator
C/Keywords: fimbria
F/1-20/Domain: signal sequence #status predicted <SIG>
F/21-151/Product: fimbria protein agfa #status experimental <MAT>

Query Match 88.1%; Score 683; DB 2; Length 151;
Best Local Similarity 90.1%; Pred. No. 1.4e-51;
Matches 136; Conservative 1; Mismatches 14; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVSGSALAGVPPWGGGNNHGGSSGPPSTLSIYOGSANAALALQ 60
DB 1 MKLLKVAFAAIVSGSALAGVPPWGGGNNHGGSSGPPSTLSIYOGSANAALALQ 60
QY 61 SDARKSETTITGSGYNGADYDQLVYTRVYTHEMAHFRNNAITDQWNAQSDITVQYCG 120
DB 61 SDARKSETTITGSGYNGADYDQVGDNSITELTONGFRNNAITDQWNAQSDITVQYCG 120
QY 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151
DB 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151

Db 121 NNAALVNOTASDSSVWVRQVGFNNATNAY 151

RESULT 2

major curlin chain precursor [imported] - Salmonella enterica subsp. enterica serovar Typhimurium
 C:Species: Salmonella enterica subsp. enterica serovar Typhimurium
 A:Note: this species has also been called Salmonella typhimurium

C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002

C:Accession: A10635

R:Parhill, J.; Dongan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, C.; Cronin, A.; Davies, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Mouton, S.; O'Garra, P.

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Nature 413, 848-852, 2001

A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar

A:Reference number: AB0502; PMID:11677608

A:Accession: A10635

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-151 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD08268.1; PID:g16502315; GSPDB:GN00176

C:Genetics:

A:Gene: STY1181

Query Match

Best Local Similarity 90.1%; Score 683; DB 2; Length 151;

Matches 136; Conservative 1; Mismatches 14; Indels 0; Gaps 0;

Db 1 MLLKVAAPAAIVVSSALAGVVPQWGGGNNHGGSGPDSITSIYQGSANALALQ 60

Db 1 MLLKVAAPAAIVVSSALAGVVPQWGGGNNHGGSGPDSITSIYQGSANALALQ 60

Qy 61 SPARKSETTITQSGYNGADYDQVLTVRVTHEMAHAFRNATIDQWNAKNSDITVQYGG 120

Db 61 SPARKSETTITQSGYNGADYDQVLTVRVTHEMAHAFRNATIDQWNAKNSDITVQYGG 120

Qy 121 NNAALVNOTASDSSVWVRQVGFNNATNAY 151

Db 121 NNAALVNOTASDSSVWVRQVGFNNATNAY 151

Qy 121 NNAALVNOTASDSSVWVRQVGFNNATNAY 151

Db 121 NNAALVNOTASDSSVWVRQVGFNNATNAY 151

Qy 121 NNAALVNOTASDSSVWVRQVGFNNATNAY 151

Db 121 NNAALVNOTASDSSVWVRQVGFNNATNAY 151

Qy 121 NNAALVNOTASDSSVWVRQVGFNNATNAY 151

Db 121 NNAALVNOTASDSSVWVRQVGFNNATNAY 151

Qy 121 NNAALVNOTASDSSVWVRQVGFNNATNAY 151

Db 121 NNAALVNOTASDSSVWVRQVGFNNATNAY 151

Qy 121 NNAALVNOTASDSSVWVRQVGFNNATNAY 151

Db 121 NNAALVNOTASDSSVWVRQVGFNNATNAY 151

A:Title: The Rpos sigma factor relieves H-NS-mediated transcriptional repression of csgA

A:Reference number: S31202; PMID:93211294; PMID:8459772

A:Accession: S31202

A:Molecule type: DNA

A:Residues: 1-6, 'V', '8-151' <OLS1>

A:Cross-references: EMBL:L04979

A:Accession: S34560

A:Molecule type: protein

A:Residues: 21-42;44-50 <OLS2>

R:Olson, A.N.; Arngvist, A.M.

submitted to the EMBL Data Library, October 1992

A:Reference number: S34559

A:Accession: S34559

A:Molecule type: DNA

A:Residues: 1-133, 'RGRDSGWLW' <OLS3>

A:Cross-references: EMBL:L04979; NID:g290424; PIDN:AAA23616.1; PID:g290425

A:Experimental source: strain K-12, substrain W3110

C:Genetics:

A:Gene: csgA

A:Map position: 23.15

C:Function:

A:Description: major component of wild-type curli; interaction between CsgA and CsgB tr

A:Note: curli are thin, coiled fibers expressed on the surface of Escherichia coli that

and H-kinogen; in the absence of CsgA, CsgB can self-assemble into polymers

F/1-20/Domain: signal sequence #status predicted <Sig>

F/21-151/Product: curlin #status experimental <Mat>

Query Match

Best Local Similarity 65.7%; Score 509; DB 2; Length 151;

Matches 103; Conservative 15; Mismatches 33; Indels 0; Gaps 0;

Db 1 MLLKVAAPAAIVVSSALAGVVPQWGGGNNHGGSGPDSITSIYQGSANALALQ 60

Db 1 MLLKVAAPAAIVVSSALAGVVPQWGGGNNHGGSGPDSITSIYQGSANALALQ 60

Qy 61 SPARKSETTITQSGYNGADYDQVLTVRVTHEMAHAFRNATIDQWNAKNSDITVQYGG 120

Db 61 SPARKSETTITQSGYNGADYDQVLTVRVTHEMAHAFRNATIDQWNAKNSDITVQYGG 120

Qy 121 NNAALVNOTASDSSVWVRQVGFNNATNAY 151

Db 121 NNAALVNOTASDSSVWVRQVGFNNATNAY 151

RESULT 4

curlin major subunit CsgA [imported] - Escherichia coli (strain O157:H7, substrain RIMD

C:Species: Escherichia coli

C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001

C:Accession: D90806

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G

gasawara, N.; Yasunaga, T.; Kohara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen

A:Reference number: A99629; PMID:11258796

A:Accession: D90806

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-152 <HAY>

A:Cross-references: GB:BA000007; PIDN:BA034843.1; PID:g13360880; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RIMD 0509952

C:Genetics:

A:Gene: Ec81420

Query Match

Best Local Similarity 62.9%; Score 487.5; DB 2; Length 152;

Matches 101; Conservative 16; Mismatches 34; Indels 1; Gaps 1;

Db 1 MLLKVAAPAAIVVSSALAGVVPQWGGGNNHGGSGPDSITSIYQGSANALALQ 59

Db 1 MLLKVAAPAAIVVSSALAGVVPQWGGGNNHGGSGPDSITSIYQGSANALALQ 60

Qy 60 QSDARKSETTITQSGYNGADYDQVLTVRVTHEMAHAFRNATIDQWNAKNSDITVQYGG 119

Db 60 QSDARKSETTITQSGYNGADYDQVLTVRVTHEMAHAFRNATIDQWNAKNSDITVQYGG 119

Qy 60 QSDARKSETTITQSGYNGADYDQVLTVRVTHEMAHAFRNATIDQWNAKNSDITVQYGG 119

Db 60 QSDARKSETTITQSGYNGADYDQVLTVRVTHEMAHAFRNATIDQWNAKNSDITVQYGG 119

Qy 60 QSDARKSETTITQSGYNGADYDQVLTVRVTHEMAHAFRNATIDQWNAKNSDITVQYGG 119

Db 60 QSDARKSETTITQSGYNGADYDQVLTVRVTHEMAHAFRNATIDQWNAKNSDITVQYGG 119

Qy 60 QSDARKSETTITQSGYNGADYDQVLTVRVTHEMAHAFRNATIDQWNAKNSDITVQYGG 119

```

Db      61 QADARNSDLTITTHGGGNGADVCGSDSSIDLTQRFNGSATIDQNGDSHTTKYQFG 120
      120 GNNAAALVNQTASDSSVWVRQVFGNNATANQY 151
      121 GGNGAAVDQTASNSTVAVTVQVFGNNATANQY 152

RESULT 5
H85665
hypothetical protein cega [imported] - Escherichia coli (strain O157:H7, substrain EDL93
C:Species: Escherichia coli
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: H85665
R:Perman, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
  Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
  Nature 409, 529-533, 2001
A>Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: H85665
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-152 <STO>
A:Cross-references: GB:AE005174; NID:G12514574; PIDN:AAG55788.1; GSPDB:GN00145; UWGP:Z16
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: cega

Query Match      62.9%; Score 487.5; DB 2; Length 152;
Best Local Similarity 66.4%; Pred. No. 7.6e-35;
Matches 101; Conservative 16; Mismatches 34; Indels 1; Gaps 1;

QY      1 MKLKVAAFAAIVVSGSALAGVPPQW-GGGGNHNGGSSGPDSTLSIYOGSANAALAL 59
      1 MKLKVAAIAIVFSSGALAGVPPQYGGGGNGGGGNGGPNSEINITYGGGNSALAL 60
      60 QSDARKSETTTTOSGYNGADYQDLVTVVTHEMAHAFRNNAITIDQNAKNSDITVGYG 119
      61 QADARNSDLTITTHGGGNGADVCGSDSSIDLTQRFNGSATIDQNGDSHTTKYQFG 120
      120 GNNAAALVNQTASDSSVWVRQVFGNNATANQY 151
      121 GGNGAAVDQTASNSTVAVTVQVFGNNATANQY 152
      Db

QY      120 GNNAAALVNQTASDSSVWVRQVFGNNATANQY 151
      121 GGNGAAVDQTASNSTVAVTVQVFGNNATANQY 152
      Db

RESULT 6
E95965
hypothetical glycine-rich protein (imported) - Sinorhizobium meliloti (strain 1021) mag
C:Species: Sinorhizobium meliloti
C>Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C:Accession: E95965
R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernat
  Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A>Title: The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing end
  A:Reference number: A95942; MUID:21396508; PMID:11481431
A:Accession: E95965
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2174 <KUR>
A:Cross-references: GB:AL591985; PIDN:CAC49389.1; PID:G15140875; GSPDB:GN00167
A:Experimental source: strain 1021, megaplasmid pSymB
R:Gilbert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
  Pella, D.; Chain, P.; Cowie, A.; Davies, R.W.; Dreano, S.; Federpiet, N.A.; Fisher, R.F.;
  L.; Hyman, R.W.; Jones, T.
  Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kise, E.; Komp, C.; Lelaure,
  hebault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yen, K.
A>Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A:Reference number: A96039; MUID:21368234; PMID:11474104
A:Contents: annotation
C:Genetics:
A:Gene: SmbD21548
A:Genome: plasmid

```

```

Query Match      13.8%; Score 107; DB 2; Length 2174;
Best Local Similarity 26.2%; Pred. No. 0.62;
Matches 37; Conservative 21; Mismatches 61; Indels 22; Gaps 6;

QY      11 AFVSGSALAGVPPQ--WGGGNGHNGGSSGPDSTLSIYOGSANAALALQSDAR---K 65
      693 AATAGAGAVGIIAGISGGGAGN--GNNATGDAFGSFGIQQGGGGGAYANTANVGRK 749
      Db

QY      66 SETTTTOSGYNGADYQDLVTVVTHEMAHAFRNNAITIDQNAK--NSDITVGYGNN 122
      750 GILTLTTQSGHAAG-----IVAGSVGGGGTGTSTASYSAGIGFTASVAVGTGNG 800
      Db

QY      123 AA-LVNQTASDSSVWVRQV 141
      801 GAGGEVSVSLTDSAIRTGCG 821
      Db

RESULT 7
S42136
cnjB protein - Tetrahymena thermophila
C:Species: Tetrahymena thermophila
C>Date: 19-Mar-1997 #sequence_revision 18-Jul-1997 #text_change 07-Dec-1999
C:Accession: S42136; S42135; S03650
R:Taylor, F.M.; Martindale, D.W.
  Submitted to the EMBL Data Library, October 1992
A:Reference number: S42136
A:Accession: S42136
A:Molecule type: DNA
A:Residues: 1-1748 <TAY>
A:Cross-references: EMBL:L03710; NID:G161751; PID:G161752
R:Taylor, F.M.; Martindale, D.W.
  Nucleic Acids Res. 21, 4610-4614, 1993
A>Title: Retroviral-type zinc fingers and glycine-rich repeats in a protein encoded by
  A:Reference number: S42135; MUID:94051569; PMID:8233798
A:Accession: S42135
A:Molecule type: DNA
A:Residues: 1164-1174; 1179-1198; 1233-1252; 1285-1293; 1297-1309; 1316-1326; 1331-1341; 1343-1
  A:Cross-references: EMBL:L03710
R:Martindale, D.W.; Taylor, F.M.
  Nucleic Acids Res. 16, 2189-2201, 1988
A>Title: Multiple introns in a conjugation-specific gene from Tetrahymena thermophila.
  A:Reference number: S03650; MUID:88189811; PMID:3357771
A:Accession: S03650
A:Molecule type: DNA
A:Residues: 236-250; '1', 252-255, 'N', 257-773 <MAR>
A:Cross-references: EMBL:X06462
C:Genetics:
A:Gene: cnjB
A:Genetic code: SGC5
A:Introns: 85/3; 136/1; 157/3; 201/2; 290/2; 327/3; 499/1; 573/2; 607/3; 708/3; 777/3; 8
  C:Keywords: zinc finger
F:1164-1450/Region: glycine-rich
F:1451-1464/Region: zinc finger CCHC motif
F:1478-1491/Region: zinc finger CCHC motif
F:1501-1514/Region: zinc finger CCHC motif
F:1530-1543/Region: zinc finger CCHC motif
F:1555-1568/Region: zinc finger CCHC motif
F:1579-1592/Region: zinc finger CCHC motif
F:1602-1615/Region: zinc finger CCHC motif
F:1626-1748/Region: glycine-rich

Query Match      12.5%; Score 97; DB 2; Length 1748;
Best Local Similarity 28.1%; Pred. No. 3.5;
Matches 36; Conservative 20; Mismatches 32; Indels 40; Gaps 7;

QY      25 QWGGGNGHNG--GNSGPDSTLSIYOGSANAALALQSDARKSETTTTOSGYNGADY 81
      1640 QGGGGNSNGCGSMGTSSGSDWN-----GQSNVQES-TTSSCGGSSGSG 1684
      Db

QY      82 DGLVTVVTHEMAHAFRNNAITIDQNAKNSDITVGYGNNAAALVNQTASDSSVWVRQV 141
      1685 NG-----TGCGWGSN--DNQOQNTGCGGWSNS--NOTNBS----- 1722
      Db

```

Qy	142	FGNNATAN	149
	:	:	:
Db	1723	WGSNNQAS	1730

RESULT 8

fimbriin proteaseGFB precursor - *Salmonella enteritidis*
C:Species: *Salmonella enteritidis*
C>Date: 31-Dec-1996 #sequence_revision 31-Dec-1996 #text_change 08-Oct-1999
C:Accession: J06040
R:Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Bansey, P.A.; Kay, W.W.
J. Bacteriol. 178, 662-667, 1996
A:Title: *Salmonella enteritidis* ageBAC operon encoding thin, aggregative fimbriae
A:Reference number: J06039; MUID:96146512; PMID:8550497
A:Accession: J06040
A:Molecule type: DNA
A:Residues: 1-151 <COL>
A:Cross-references: GB:U43280, NID:g1184712, PIDN:AAC43598.1, PID:g1184713
A:Experimental source: strain 276755-3b
C:Genetics:
A:Gene: agfB
C:Function:
A:Description: minor component of thin aggregative fimbriae
A>Note: fimbriae bind to fibronectin, plasminogen, tissue plasminogen activator
C:Keywords: fimbria
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-151/Product: fimbriin protein agfB #status predicted <MAT>

Query Match	12.2%	Score 94.5	DB 2	length 151
Best Local Similarity	26.9%	Pred. No. 0.34		
Matches 35; Conservative	17;	Mismatches	41;	Gaps 4;

```

QY 10 AATVVGSSALAGVPPMGCGGNNHNGGNSGGDSTLTSTYYCGANMALLGDSARKSETT 69
   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 58 ARVRGSGSKLTSTVISC-EGGNRRAKRVDGAGYNNRATVTEGTGMAN-----DAS 109
   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY 70 ITTSGTGNCGADYQTLVTRVYVTHMMAAFNNNTITDGNNAKSDITVGYCGNNMALLNQOT 129
   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 104 ISQSAIGNSA-----AIIKSGGNNKNNIT--GYGKTAKVAVVQK 140

```

```
QY      130 ASDSSVMVRQ 139
          | :: | |
Db      141 QSHMAIRVTQ 150
```

RESULT 9

Probable PPE protein - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 15-Sep-2003
C:Accession: F70825
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skellton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrett, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:96295987; PMID:9634230
A:Accession: F70825
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-645 <COL>
A:Cross-references: GB:AL021958; GB:AL123456; NID:g3261536; PIDN:CAM17522.1; PID:e125322
A:Experimental source: strain H37Rv
C:Genetics:
C:Gene: PPE

Query Match	12.1%	Score 94	DB 2	Length 645
Best Local Similarity	24.3%	Pred. NO. 2		
Matches 34	Conservative 14	Mismatches 42	Indels 50	Gaps 7

```
QY      11 AIVWGSALAGVPEQMGGGGHN-GGNSSPDSTLIYQYSANALLAQDPARKSETT   69
          | : | : | ||| | ||| :
Db     294 AVPTPGNGNVGI-----GNGNGNFGGGNT-----GNANIGINVGDG-----   333
```

QY 70 ITSGCYGNAGNDYDOLVTRVYTHEMAHAFRNATIDQNAKSDITVQCYGGNNAALVNOT 122

DB 332 --NVGFGNGSGSYN-----FGFGNTG-----NNNIGIGITGSGNQIGFGLN 366

RESULT 10

1.Leishmanolysin (EC 3.4.24.36) precursor [validated] - Leishmania major
 N,Alternate names: promastigote surface proteinase; surface endopeptidase
 C,Species: Leishmania major
 C,Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 15-Sep-2000
 C,Accession: P10221; A27598; A60648
 R,Button, L.L.; McMaster, W.R.
 J. Exp. Med. 171, 589, 1990
 A,Reference number: P10221
 A,Contents: erratum
 A,Accession: P10221
 A,Molecule type: DNA
 A,Residues: 1-602 <BUT>
 A,Cross-references: GB:V00647; NID:g9554; PIDN:CA8673.1; PID:g9555
 A,Note: this is a revision to the sequence from reference A27598
 R,Button, L.L.; McMaster, W.R.
 J. Exp. Med. 167, 724-729, 1988
 A,Title: Molecular cloning of the major surface antigen of Leishmania.
 A,Reference number: A27598; MUID:88154764; PID:3346625
 A,Accession: A27598

A:Status: significant sequence differences
A:Molecule type: DNA
R,Bouvier, J.; Bordier, C.; Vogel, H.; Reichelt, R.; Etges, R.
Mol. Biochem. Parasitol. 37, 235-246, 1989
A:Title: Characterization of the promastigote surface protease of *Leishmania* as a member
A:Accession number: A60648; MUID:90114330; PMID:2608099
A:Accession: A60648
A:Molecule type: protein
A:Residues: 101,'E',103-118,'SV',121-123 <BOU>
A:Experimental source: strain LBMS13
R,Schlegelmant, E.; Etges, R.; Metcalf, P.
Submitted to the Brookhaven Protein Data Bank, March 1997.

A,Description: catalyzes the hydrolysis of peptide bonds between two hydrophobic residues
A,Note: the activated form can activate the proenzyme form
C,Superfamily: leishmanolysin
C,Keywords: blocked carboxyl end; cell adhesion; glycoprotein; homodimer; hydrolase; li
F,1-39/Domain: signal sequence #status predicted <SIG>
F,40-100/Domain: activation peptide #status predicted <AP>
F,101-577/Product: leishmanolysin #status experimental <MP>
F,578-602/Domain: carboxyl-terminal propeptide #status predicted <CP>
F,48,264,268,334/Binding site: zinc, catalytic (Cys, His, His, His) (inhibited) #status
F,100,101/Cleaveage site: Val1-Val1 (autolytic) #status experimental
F,125-142,191-230,314-386,393-455,406-425,415-489,466-510,515-565,535-558/Disulfide bon
F,245,268,334/Binding site: zinc, catalytic (His) (active) #status experimental
F,265/Active site: Glu #status predicted
F,300,407/Binding site: carboxylate (Asn) (covalent) #status experimental
F,577/Modified site: GPI-anchor ethanolamine amidated carboxyl end (Asn) (in mature form

```

Query Match      11.8%; Score 91.5; DB 1; Length 602;
Best Local Similarity 41.7%; Pred. No. 3;
Matches 25; Conservative 2; Mismatches 12; Indels 21; Gaps 2
QY      81 YDGLVTRVYTHMAHAF-----RNNATIQWAKNSDIV-----GQYG 119
          |||||
          | | | | |

```

Db 254 YDOLVTRVVTHEMAHALGFSGFEDARIVANPNVRGNKFDVPVINSSTAVAKAREQYX 313

RESULT 11

C86266

F3F19.21 protein - Arabidopsis thaliana

C/Species: Arabidopsis thaliana (mouse-ear cress)

C/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Nov-2001

C/Accession: C86266

R/Theologian: A.; Eckert, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;

Ansen, N.F.; Hughes, B.; Hultzer, L.

Nature 408, 816-820, 2000

A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A/Reference number: A86141; MUID:21016719; PMID:11130712

A/Accession: C86266

A/Status: Preliminary

A/Molecule type: DNA

A/Residues: 1-573 <STO>

A/Cross-references: GB:AE005172; NID:g4850402; PIDN:AAD31072.1; GSPDB:GN00141

C/Genetic: 1

A/Map position: 1

Query Match

Best Local Similarity 11.7%; Score 91; DB 2; Length 573;

Matches 36; Conservative 19; Mismatches 49; Indels 30; Gaps 7;

QY 13 VWSGALAGVPMWGGGNGNNGSGPDSSTLSIYOGSANAALALQSDARKSETTTTQ 72

Db 84 IVGGGTVEG---KYRNDGGHNG---ISGPTSDVDYPOQSSPAKGLNIDIQNK---IQ 135

QY 73 SGYGNADIDQVLTTRVVTHEMAHAFRNA-TIDQNAKNS-----DITYGYGGNN 122

Db 136 OG-----STTVLVNN--HGFGNAVNVPEMVPVHNSYGAPPOGAQOIPVSOQSVNP 183

QY 123 AALVNOTASDSSVM 136

Db 184 NVWNAKSPQSFV 197

RESULT 12

E70663

Probable PPE protein - Mycobacterium tuberculosis (strain H37RV)

C/Species: Mycobacterium tuberculosis

C/Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 15-Sep-2003

C/Accession: E70663

R/Coile, S.T.; Brosch, R.; Parhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A/Authors: Sgarbes, R.; Sultson, J.E.; Taylor, K.; Whitehead, S.; Bartell, B.G.

A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A/Reference number: A70500; MUID:98255981; PMID:9634230

A/Accession: E70663

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-615 <COL>

A/Cross-references: GB:Z83860; GB:AL123456; NID:g3261681; PIDN:CAB06165.1; PID:e290763;

C/Genetic: 1

A/Experimental source: strain H37RV

A/Map position: 1

Query Match

Best Local Similarity 11.7%; Score 91; DB 2; Length 615;

Matches 31; Conservative 14; Mismatches 47; Indels 46; Gaps 5;

QY 15 SGGALAGVPMWGGGNGNNGSGPDSSTLSIYOGSANAALALQSDARKSETTTTQ 74

Db 11.7%; Score 91; DB 2; Length 615;

Matches 31; Conservative 14; Mismatches 47; Indels 46; Gaps 5;

QY 15 SGGALAGVPMWGGGNGNNGSGPDSSTLSIYOGSANAALALQSDARKSETTTTQ 74

Db 11.7%; Score 91; DB 2; Length 615;

Db 365 SGTGNGIG----FGNGNNNGFNGSG-DGNIFRNSGDN-----TG 401

QY 75 YENGADYDQVLTTRVVTHEMAHAFRNAATIDQNAKNSDITVYOGSANAALV---NQTS 131

Db 402 FGNA-----GNITGFGNAGNLNTGFSAGNGNVIRDGSGNSNG 441

QY 132 DSSVAVROVFGNNATAN 149

Db 442 SFNVGFQNTGFGNSAGN 459

RESULT 13

S70787

Culin nucleator protein csbB precursor - Escherichia coli (strain K-12)

N/Alternate names: csbB protein; curlin nucleation component; minor curlin protein

C/Species: Escherichia coli

C/Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 01-Mar-2002

C/Accession: S70787; F64846

R/Hammart, M.; Arnyavast, A.; Bian, Z.; Olsen, A.; Normark, S.

Mol. Microbiol. 18, 661-670, 1995

A/Title: Expression of two csb operons is required for production of fibronectin- and C

A/Reference number: S70783; MUID:96414468; PMID:8817489

A/Accession: S70787

A/Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-151 <HAM>

A/Cross-references: EMBL:X90754; NID:g1147558; PIDN:CMA62281.1; PID:g1147563

A/Experimental source: strain K12, substrain W3110

A/Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995

R/Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C

.A.; Rose, D.J.; Mau, B.; Shaio, Y.

Science 277, 1453-1462, 1997

A/Title: The complete genome sequence of Escherichia coli K-12.

A/Reference number: A64720; MUID:97426617; PMID:9278503

A/Accession: F64846

A/Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-151 <BLAT>

A/Cross-references: GB:AE000205; GB:U00096; NID:g1787265; PIDN:AC74125.1; PID:g1787278,

A/Experimental source: strain K-12, substrain MG1655

C/Genetic: 1

A/Map position: 23.15

A/Description: minor component of wild-type curli; interaction between CsgA and CsgB tri

A/Note: curli are thin, coiled fibers expressed on the surface of Escherichia coli that

F/1-21/Domain: signal sequence #status predicted <SIG>

F/22-151/Product: minor curlin chain #status predicted <MAT>

Query Match

Best Local Similarity 11.5%; Score 89.5; DB 2; Length 151;

Matches 32; Conservative 15; Mismatches 62; Indels 5; Gaps 2;

QY 38 SSGPDSSTLSIYOGSANAALALQSDARKSETTTTQSGNGADYDQVLTTRVVTHEMAHAF 97

Db 21 AAGYDLANSEYVF-----AVNELSKSSFQAALIGAGTNNSHQLQSGSKLAAVVAQEGS 76

QY 98 RNNATIDQNAKNSDITVYOGSANAALVNCASSSVWVWVQVGGNNATANOY 151

Db 77 SNRAKIDQGDYNL-AYIDQASANDASISQAGYNTAMTIQKSGNKNANTQY 129

RESULT 14

C90806

minor curlin subunit precursor CsgB (imported) - Escherichia coli (strain O157:H7, subsp

C/Species: Escherichia coli

C/Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001

C/Accession: C90806

R/Hayashi, T.; Makino, K.; Onishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.

gasekara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hatcort, M.; Shingawa, H.

DNA Res. 8, 11-22, 2001

A/Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene

A:Reference number: A99629; MUID:21156231; PMID:11258796
 A:Accession: C90806
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-151 <HAY>
 A:Cross-references: GB:BA000007; PIDN:BA834842.1; PID:Q13360879; GSPDB:GN00154
 A:Experimental source: strain O157:H7, substrain R1MD 0509952
 C:Genetics:
 A:Gene: ECs1419

Query Match 11.5%; Score 89.5; DB 2; Length 151;
 Best Local Similarity 28.1%; Pred. No. 0.92;
 Matches 32; Conservative 15; Mismatches 62; Indels 5; Gaps 2;

QY 38 SSGPDSTISTYQYGSAALALQSDARKSETTITQSGYGAGADYDQLVTRVTHEMAHAF 97
 DB 21 AAGYDLANSEYNF---AVNELSKSSFNOAALIGQAGTNSAQLRQGGSKLAVVAQEGS 76

QY 98 RNNATIDQWNAKNSDITVQYGGNNAALVNOTASDSVWVRQVGFNNATANOY 151
 DB 77 SNRAKIDQTDGYNL-AYIDQAGSANDASISQAGYGTMTMTIQKSGNKANITQY 129

RESULT 15

G85665
 curlin minor chain precursor, CsgA homolog [imported] - Escherichia coli (strain O157:H7)
 C:Species: Escherichia coli
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 C:Accession: G85665
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
 Miller, L.; Grobbeck, E.J.; Davis, N.W.; Lam, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
 Nature 409, 529-533, 2001
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A:Reference number: A85480; MUID:21074935; PMID:11206551
 A:Accession: G85665
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-151 <STO>
 A:Cross-references: GB:AE005174; NID:G12514573; PIDN:AG55787.1; GSPDB:GN00145; UMG:P:Z16
 A:Experimental source: strain O157:H7, substrain EDL933
 C:Genetics:
 A:Gene: csgA

Query Match 11.5%; Score 89.5; DB 2; Length 151;
 Best Local Similarity 28.1%; Pred. No. 0.92;
 Matches 32; Conservative 15; Mismatches 62; Indels 5; Gaps 2;

QY 38 SSGPDSTISTYQYGSAALALQSDARKSETTITQSGYGAGADYDQLVTRVTHEMAHAF 97
 DB 21 AAGYDLANSEYNF---AVNELSKSSFNOAALIGQAGTNSAQLRQGGSKLAVVAQEGS 76

QY 98 RNNATIDQWNAKNSDITVQYGGNNAALVNOTASDSVWVRQVGFNNATANOY 151
 DB 77 SNRAKIDQTDGYNL-AYIDQAGSANDASISQAGYGTMTMTIQKSGNKANITQY 129

Search completed: March 11, 2004, 18:42:12
 Job time : 11.3 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:14:48 ; Search time 6.3 Seconds

(without alignments)
1248.031 Million cell updates/sec

Title: US-09-543-407-28

Perfect score: 775
Sequence: 1 MKLLKVAFAFAIVSGSALA.....DSSVMYRQVGFNNATPANOY 151

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	ID	Description
1	68.1	88.1	151	1	CSGA_SALTY
2	50.9	65.7	151	1	CSGA_SALTY
3	487.5	62.9	152	1	CSGA_SALTY
4	94.5	12.2	151	1	CSGA_SALTY
5	91.5	11.8	602	1	GP63_SALTY
6	89.5	11.5	151	1	CSGA_SALTY
7	89.5	11.5	1322	1	ICEN_PANAN
8	89.5	11.5	1567	1	ICEN_XANCT
9	87.5	11.3	1034	1	ICEN_PANAN
10	87.5	11.3	1258	1	ICEN_ERWHE
11	86.5	11.2	151	1	CSGA_SALTY
12	86.5	11.1	590	1	GP63_SALTY
13	84.5	10.9	646	1	GP63_SALTY
14	84.5	10.8	599	1	GP63_SALTY
15	84.5	10.8	1608	1	ALBYA_SERMA
16	83.5	10.8	856	1	ALBYA_SERMA
17	83.5	10.7	1656	1	OMP8_RICJA
18	81.5	10.5	392	1	HMB1_HUMAN
19	81.5	10.5	697	1	SIM_DROME
20	81.5	10.5	1028	1	OV0_DROME
21	80.5	10.3	331	1	OMB2_NEHMB
22	80.5	10.3	760	1	YBIL_ECOLI
23	79.5	10.3	475	1	ABPB_STRCO
24	79.5	10.3	548	1	CEAK_ECOLI
25	79.5	10.3	1210	1	ICEN_PSEFL
26	79.5	10.2	678	1	YF48_MYCTU
27	79.5	10.2	823	1	DHN3_YEAST
28	78.5	10.1	232	1	DHN3_PEA
29	77.5	10.0	1063	1	IMD_ARTCO
30	77.5	10.0	1641	1	SPTS_YEAST
31	77.5	10.0	1148	1	ICEN_PSEFL
32	77.5	10.0	1185	1	MAPX_DROME
33	77.5	10.0	1196	1	ICEV_PSEFL

34	77	9.9	204	1	CORA_MEDSA
35	77	9.9	376	1	P33_MYCPB
36	77	9.9	439	1	ACMA_LACLA
37	76.5	9.9	487	1	Y442_MYCTU
38	76.5	9.9	881	1	YFCU_ECOLI
39	76	9.8	147	1	HPAA_CAUCR
40	76	9.8	362	1	P35_MYCPB
41	76	9.8	542	1	SCMB_YEAST
42	76	9.8	1571	1	C3G_DROME
43	75.5	9.7	163	1	HCV_NATPH
44	75.5	9.7	394	1	OMG1_SALTY
45	75.5	9.7	424	1	COAA_BPFD

ALIGNMENTS

RESULT 1	CSGA_SALTY	STANDARD	PRT	151 AA
AC	P55225			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Major curlin subunit precursor (Fimbriae SFR17).			
GN	CSGA OR AGFA OR STM1144 OR STY1181 OR T11776.			
OS	Salmonella typhimurium			
OS	Salmonella enteritidis			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OC	Enterobacteriaceae; Salmonella.			
OX	NCBI_TaxID=602, 601, 592;			
LN	(1)			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=S. typhimurium; STRAIN=SR-11;			
RA	MEDLINE=98117058; PubMed=9457880;			
RT	Romling U., Bian Z., Hammar M., Sierralta W.D., Normark S.;			
RT	"Curli fibers are highly conserved between Salmonella typhimurium and			
RT	Escherichia coli with respect to operon structure and regulation.";			
RL	J. Bacteriol. 180:722-731(1998).			
LN	(2)			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=S. typhimurium; STRAIN=LT2 / SGGC1412 / ATCC 700720;			
RA	MEDLINE=21534948; PubMed=11677609;			
RA	McClelland M., Sanderson K.E., Speleth J., Clifton S.W., Latreille P.,			
RA	Courney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,			
RA	Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,			
RA	Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,			
RA	Waterston R., Wilson R.K.;			
RT	"Complete genome sequence of Salmonella enterica serovar Typhimurium			
RT	LT2.";			
RL	Nature 413:852-856(2001).			
LN	(3)			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=S. typhimurium; STRAIN=CT18;			
RA	MEDLINE=21534947; PubMed=11677608;			
RA	Parthill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,			
RA	Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebatina M.,			
RA	Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,			
RA	Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrer J.,			
RA	Fellwell T., Hamlin N., Haque A., Hien T.T., Holtroyd S., Jagels K.,			
RA	Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,			
RA	Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,			
RA	Whithead S., Barrett B.G.;			
RT	"Complete genome sequence of a multiple drug resistant Salmonella			
RT	enterica serovar Typhimurium CT18.";			
RL	Nature 413:848-852(2001).			
LN	(4)			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=S. typhimurium; STRAIN=Ty2 / ATCC 700931;			
RA	MEDLINE=22531167; PubMed=12644504;			
RA	Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,			
RA	Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;			

RT "Comparative genomics of *Salmonella enterica* serovar Typhi strains Ty2
 RL and C18.";
 RL J. Bacteriol. 185:2330-2337(2003).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S.enteritidis; STRAIN=27655-3B;
 RX MEDLINE=96146512; PubMed=8550497;
 RA Collinson S.K., Clouthier S.C., Doran J.L., Baner P.A., Kay W.W.;
 RT "Salmonella enteritidis agfBAC operon encoding thin, aggregative
 RT fimbriae.";
 RL J. Bacteriol. 178:662-667(1996).
 RN [6]
 RP SEQUENCE OF 21-151 FROM N.A.
 RC SPECIES=S.enteritidis; STRAIN=27655-3B;
 RX MEDLINE=94013373; PubMed=8104955;
 RA Doran J.L., Collinson S.K., Burian J., Santos G., Todd E.C.D.,
 RA Munro C.K., Kay C.M., Baner P.A., Peterkin P.I., Kay W.W.;
 RT "DNA-based diagnostic tests for *Salmonella* species targeting agfA,
 RT the structural gene for thin, aggregative fimbriae.";
 RL J. Clin. Microbiol. 31:2263-2273(1993).
 RN [7]
 RP SEQUENCE OF 21-33.
 RC SPECIES=S.enteritidis; STRAIN=27655-3B;
 RX MEDLINE=91310586; PubMed=1677357;
 RA Collinson S.K., Emsedy L., Mueller K.-M., Trust T.J., Kay W.W.;
 RT "Purification and characterization of thin, aggregative fimbriae from
 RT *Salmonella enteritidis*.";
 RL J. Bacteriol. 173:4773-4781(1991).
 CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 CC FIBRONECTIN.
 CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AJ002301; CAA05317.1; -
 DR EMBL; AE008749; AAL20074.1; -
 DR EMBL; AL627263; CAD08268.1; -
 DR EMBL; AF016840; AAO63399.1; -
 DR EMBL; U03280; AAC33599.1; -
 DR PIR; J06039; J06039.
 DR StyGene; SG10608; csGA.
 DR Fimbrin; Signal; Complete proteome.
 KW Fimbrin; Signal; Complete proteome.
 FT CHAIN 1 20
 FT SIGNAL 21 151
 FT CONFLICT 134 151 SVMWQVGFNNATANOY -> DSYTOVAS (IN
 FT REF. 6)
 SQ SEQUENCE 151 AA; 15305 MW; B7DAC0D16B621359 CRC64;
 Query Match 88.1%; Score 683; DB 1; Length 151;
 Best Local Similarity 90.1%; Pred. NO. 1.3e-51;
 Matches 136; Conservative 1; Mismatches 14; Indels 0; Gaps 0;
 Oy 1 MKLKVAAFAAIVVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANAAALALQ 60
 Db 1 MKLKVAAFAAIVVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANAAALALQ 60
 Oy 61 SDARKSETTTTQSGYGNADYDQLVTRVTHMAHAFRNATTDQWAKNSDITVGGYGG 120
 Db 61 SDARKSETTTTQSGYGNADYDQGVGADNSTIELTNGCFRNATTDQWAKNSDITVGGYGG 120
 Oy 121 NNAALVNOTASDSSVWVRQVGFNNATANOY 151
 Db 121 NNAALVNOTASDSSVWVRQVGFNNATANOY 151

RESULT 2
 ID CSGA_ECOLI STANDARD; PRT; 151 AA.
 AC P28307.
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Major curlin subunit precursor.
 GN CSGA OR B1042.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / W3110.
 RX MEDLINE=93211294; PubMed=8459772;
 RA Olsen A., Arngvist A.;
 RT "The RpoS sigma factor relieves H-NS-mediated transcriptional
 RT repression of csGA, the subunit gene of fibronectin-binding curli in
 RT *Escherichia coli*.";
 RL Mol. Microbiol. 7:523-536(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MC4100;
 RX MEDLINE=96414468; PubMed=8817489;
 RA Hammar M., Arngvist A., Bian Z., Olsen A., Normark S.;
 RT "Expression of two csG operons is required for production of
 RT fibronectin- and congo red-binding curli polymers in *Escherichia coli*
 RT K-12.";
 RL Mol. Microbiol. 18:661-670(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blatner F.R., Plunkett G., III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of *Escherichia coli* K-12.";
 RL Science 277:1453-1474(1997).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97061202; PubMed=8905232;
 RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
 RA Mori H., Kotonura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
 RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horinchi T.;
 RT "A 718-bp DNA sequence of the *Escherichia coli* K-12 genome
 RT corresponding to the 12.7-28.0 min region on the linkage map.";
 RL DNA Res. 3:137-155(1996).
 RN [5]
 RP SEQUENCE OF 21-40.
 RC STRAIN=K12 / YMEB;
 RX MEDLINE=93023873; PubMed=1357528;
 RA Arngvist A., Olsen A., Pfeiffer J., Russell D.G., Normark S.;
 RT "The Crl protein activates cryptic genes for curli formation and
 RT fibronectin binding in *Escherichia coli* HB101.";
 RL Mol. Microbiol. 6:2443-2452(1992).
 RN [6]
 RP SEQUENCE OF 21-31.
 RC MEDLINE=91310586; PubMed=1677357;
 RA Collinson S.K., Emsedy L., Trust T.J., Kay W.W.;
 RT "Purification and characterization of thin, aggregative fimbriae from
 RT *Salmonella enteritidis*.";
 RL J. Bacteriol. 173:4773-4781(1991).
 CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 CC FIBRONECTIN.


```

CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L04979; AAA23616.1; -
DR EMBL; X90754; CA62282.1; -
DR EMBL; AE000205; AAC7426.1; -
DR EMBL; D90741; BAA35832.1; -
DR EMBL; D90742; BAA35840.1; -
DR PIR; S70788; S70788.
DR Ecogen; EG11489; CSGA.
DR Fimbrtia; Signal; Complete proteome.
FT SIGNAL 1 20
FT CHAIN 21 151 MAJOR CURLIN SUBUNIT.
FT CONFLICT 7 7 A -> B (IN REF. 1).
SQ SEQUENCE 151 AA; 15049 MW; C003470D208D395F CRC64;

Query Match 65.7%; Score 509; DB 1; Length 151;
Best Local Similarity 68.2%; Pred. No. 8.7e-37;
Matches 103; Conservative 15; Mismatches 33; Indels 0; Gaps 0;

QY 1 MLKLKVAAPFAIVGSGALAGVPPWQGGGNNHNGSGSPDSTLSIYQGSANALAL 60
DB 1 MLKLKVAALTAIVFSSGALAGVPPQYGGGNNHNGSGNNSPNELNIYQYGGNSALAL 60

QY 61 SDARKSETTITQSGYNGADVDQLVTRVVTHEMAFPRNNATTIDQWAKNSDITVQYGG 120
DB 61 TDARNSDLTITQHGCGAGADVQGSDDSDSLTQRFNGSATLQDQNGKSEMTVQFQGG 120

QY 121 NNAALVNOTASDSSVMRQVGFNNATPANOY 151
DB 121 GNGAAVDQTASNSVTNVQVGFNNATPANOY 151

RESULT 3
CSGA_ECO57 STANDARD; PRT; 152 AA.
ID CSGA_ECO57
AC Q93024;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Major curlin subunit precursor.
GN CSGA OR Z1676 OR ECS1420.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_Taxid=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / ATCC 43895;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grodeck E.J., Davis N.W., Lim A., Dimailanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
RT Nature 409:529-533(2001).

```

```

RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Toke T.,
RA Iida T., Takami H., Honda T., Hattori M., Shinagawa H.,
RA Kuhara S., Shiba T., Hattori M., Sasaki C., Ogasawara N., Yasunaga T.,
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12."
RL DNA Res. 8:11-22(2001).

CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLIN. CURLIN ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLIN CAN BIND TO
CC FIBRONECTIN.
CC -----
CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF275733; AAK53212.1; -
DR EMBL; AE005315; AAG55788.1; -
DR EMBL; AP002554; BAB34843.1; -
DR PIR; D90806; D90806.
DR PIR; H85665; H85665.
DR Fimbrtia; Signal; Complete proteome.
FT SIGNAL 1 20
FT CHAIN 21 152 MAJOR CURLIN SUBUNIT.
SQ SEQUENCE 152 AA; 15099 MW; EE2D2D94DDE91243 CRC64;

Query Match 62.9%; Score 487.5; DB 1; Length 152;
Best Local Similarity 66.4%; Pred. No. 6e-35;
Matches 101; Conservative 16; Mismatches 34; Indels 1; Gaps 1;

QY 1 MLKLKVAAPFAIVGSGALAGVPPWQGGGNNHNGSGSPDSTLSIYQGSANALAL 59
DB 1 MLKLKVAALTAIVFSSGALAGVPPQYGGGNNHNGSGNNSPNELNIYQYGGNSALAL 60

QY 60 QSDARKSETTITQSGYNGADVDQLVTRVVTHEMAFPRNNATTIDQWAKNSDITVQYGG 119
DB 61 QADARNSDLTITQHGCGAGADVQGSDDSDSLTQRFNGSATLQDQNGKSDHMTVQFQGG 120

QY 120 GNAALVNOTASDSSVMRQVGFNNATPANOY 151
DB 121 GNGAAVDQTASNSVTNVQVGFNNATPANOY 152

RESULT 4
CSGB_SALT1 STANDARD; PRT; 151 AA.
ID CSGB_SALT1
AC P55226;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE Minor curlin subunit precursor (Fimbrin Sef17 minor subunit).
GN CSGB OR AGFB OR STM1143.
OS Salmonella typhimurium, and
OS Salmonella enteritidis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_Taxid=602, 592;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S. typhimurium; STRAIN=SR-11;
RX MEDLINE=98117058; PubMed=9457880;
RA Rowling U., Bian Z., Hammar M., Sierralta W.D., Normark S.;
RT "Curli fibers are highly conserved between Salmonella typhimurium and
RT Escherichia coli with respect to operon structure and regulation."

```

```

RN J. Bacteriol. 180:722-731(1998).
RN (2)
RN SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=LT2 / SGC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Bowtell S., Ali J., Dante M., Du P., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Milvanev E.,
RA Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RA "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2."
RL Nature 413:852-856(2001).
RN (3)
RN SEQUENCE FROM N.A.
RC SPECIES=S.enteritidis; STRAIN=27655-3B;
RX MEDLINE=96146512; PubMed=8550497;
RA Collinson S.K., Clouthier S.C., Doran J.L., Baner P.A., Kay W.W.;
RA "Salmonella enteritidis ageBAC operon encoding thin, aggregative
RT fimbriae."
RL J. Bacteriol. 178:662-667(1996).
CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREPREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
CC CURLIN MONOMERS.
CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, AJ002301; CA05316.1; -
DR EMBL, AE008749; AAL20073.1; -
DR EMBL, U43280; AAC43598.1; -
DR PIR, JC6040; JC6040.
DR StyGene; SG10609; CSGB.
KW Fimbria; Signal; Complete proteome.
FT SIGNAL
FT CHAIN 1 21 POTENTIAL.
FT SEQUENCE 151 AA; 16182 MW; C0FC5430B6D361D CRC64;
SQ
Query Match 12.2%; Score 94.5; DB 1; Length 151;
Best Local Similarity 26.9%; Pred. No. 0.19;
Matches 35; Conservative 17; Mismatches 41; Indels 37; Gaps 4;
OY 10 AAIYVSGALAGVVPQWGGNGNNGNSGPDSTLSIYQYGSNAAALALQSDARKSETT 69
DB 58 ARVQEGSGKLSVLSQ--EGGNRAKVDQAGNYFAIEGNGN-----DAS 103
OY 70 ITGSGYGAGDYDQLVTRVTHMAHAFRNNAITDQNNAKKSDITVQYGSNAAALVNOT 129
DB 104 ISQARVGSNA-----AIIQKSGNKANIT--QYGTQRTAAVAVVOK 140
OY 130 ASDGSVMVRQ 139
DB 141 QSHMAIRVTO 150

```

```

GN Gp63.
OS Leishmania major.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN (1)
RN SEQUENCE FROM N.A., AND SEQUENCE OF 101-123.
RX MEDLINE=88154764; PubMed=3346625;
RA Button L.L., McMaster W.R.;
RT "Molecular cloning of the major surface antigen of Leishmania."
RL J. Exp. Med. 167:724-729(1988).
RN (2)
RN REVISIONS.
RA Button L.L., McMaster W.R.;
RL J. Exp. Med. 171:589-589(1990).
RN (3)
RN GPI-ANCHOR.
RX MEDLINE=91009116; PubMed=2145267;
RA Schneider P., Ferguson M.A.J., McConville M.J., Mehlex A.,
RA Homans S.W., Border C.;
RT "Structure of the glycosyl-phosphatidylinositol membrane anchor of
RL the Leishmania major promastigote surface protease."
RN J. Biol. Chem. 265:16955-16964(1990).
RN (4)
RN X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).
RX MEDLINE=95406217; PubMed=7675788;
RA Schlegelhauf E., Etges R., Metcalf P.;
RT "Crystallization and preliminary X-ray diffraction studies of
RT Leishmania major surface metalloproteinase from Leishmania
RT major."
RL Proteins 22:58-66(1995).
RN (5)
RN X-RAY CRYSTALLOGRAPHY (1.86 ANGSTROMS).
RX MEDLINE=98416698; PubMed=9739094;
RA Schlegelhauf E., Etges R., Metcalf P.;
RT "The crystal structure of the Leishmania major surface proteinase
RT Leishmanolysin."
RL Structure 6:1035-1046(1998).
CC -1- FUNCTION: Has an integral role during the infection of macrophages
CC in the mammalian host.
CC -1- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and
CC P1', and basic residues at P2 and P3'. A model nonapeptide is
CC cleaved at -Ala-Tyr-[Leu-Lys-Lys-.
CC -1- COFACTOR: Binds 1 zinc ion per subunit.
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -1- PTM: THE PHOSPHATIDYLINOSITOL MOIETY OF THE GPI-ANCHOR CONTAINS A
CC FULLY SATURATED, UNBRANCHED 1-O-ALKYL CHAIN (MAINLY C24:0) AND A
CC MIXTURE OF FULLY SATURATED UNBRANCHED 2-O-ACYL CHAINS (C12:0,
CC C14:0, C16:0, AND C18:0).
CC -1- SIMILARITY: Belongs to peptidase family M8.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, Y00647; CA068673.1; -
DR PIR, P10221; P10221.
DR PDB, 1LMU; 17-SEP-97.
DR MEROPS, M08.001; -.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR InterPro; IPR001577; Peptidase_M8.
DR Pfam; PF01457; Peptidase_M8; 1.
DR PRINTS; PR00782; LSHMANOLYSIN.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydrolase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;
KW Zymogen; Signal; Cell adhesion; GPI-anchor; 3D-structure; Lipoprotein.
FT SIGNAL 1 39 POTENTIAL.
FT PROPEP 40 100 ACTIVATION PEPTIDE.
FT CHAIN 101 577 LEISHMANOLYSIN.
FT PROPEP 578 602 REMOVED IN MATURE FORM.

```

FT	METAL	264	ZINC (CATALYTIC).
FT	ACT SITE	264	
FT	METAL	265	ZINC (CATALYTIC).
FT	METAL	268	ZINC (CATALYTIC).
FT	DISULFID	334	
FT	DISULFID	125	
FT	DISULFID	191	
FT	DISULFID	125	
FT	DISULFID	314	
FT	DISULFID	393	
FT	DISULFID	406	
FT	DISULFID	445	
FT	DISULFID	415	
FT	DISULFID	466	
FT	DISULFID	510	
FT	DISULFID	515	
FT	DISULFID	535	
FT	CARBOHYD	300	
FT	CARBOHYD	407	
FT	LIPID	577	
FT	STRAND	101	
FT	STRAND	102	
FT	STRAND	107	
FT	STRAND	111	
FT	HELIX	116	
FT	TURN	121	
FT	TURN	128	
FT	STRAND	131	
FT	STRAND	133	
FT	STRAND	139	
FT	STRAND	141	
FT	HELIX	144	
FT	HELIX	150	
FT	TURN	159	
FT	HELIX	160	
FT	TURN	169	
FT	STRAND	170	
FT	STRAND	171	
FT	STRAND	172	
FT	STRAND	174	
FT	STRAND	177	
FT	STRAND	180	
FT	TURN	189	
FT	HELIX	191	
FT	HELIX	193	
FT	TURN	202	
FT	TURN	203	
FT	STRAND	205	
FT	STRAND	210	
FT	TURN	221	
FT	STRAND	226	
FT	TURN	232	
FT	STRAND	234	
FT	STRAND	238	
FT	HELIX	244	
FT	HELIX	249	
FT	HELIX	256	
FT	TURN	270	
FT	HELIX	271	
FT	TURN	274	
FT	TURN	280	
FT	STRAND	283	
FT	HELIX	289	
FT	HELIX	296	
FT	HELIX	302	
FT	TURN	313	
FT	TURN	315	
FT	STRAND	320	
FT	TURN	328	
FT	STRAND	332	
FT	STRAND	335	
FT	TURN	337	
FT	TURN	341	
FT	STRAND	343	
FT	TURN	344	
FT	STRAND	353	
FT	STRAND	353	
FT	HELIX	364	
FT	TURN	365	
FT	STRAND	369	
FT	HELIX	372	
FT	TURN	374	
FT	TURN	380	
FT	HELIX	386	
FT	STRAND	394	
FT	TURN	395	
FT	STRAND	397	
FT	STRAND	398	
FT	TURN	402	
FT	STRAND	413	
FT	TURN	414	
FT	TURN	417	

N-LINKED (GLCNAC. . .) (POTENTIAL).
N-LINKED (GLCNAC. . .) (POTENTIAL).
GPI-anchor amidated asparagine.

FT	STRAND	421	425
FT	STRAND	428	429
FT	HELIX	435	437
FT	TURN	443	444
FT	STRAND	445	446
FT	TURN	450	454
FT	STRAND	458	465
FT	TURN	466	467
FT	HELIX	470	472
FT	TURN	475	477
FT	HELIX	478	480
FT	TURN	485	486
FT	STRAND	487	494
FT	STRAND	496	496
FT	STRAND	506	516
FT	TURN	517	520
FT	TURN	521	525
FT	TURN	527	528
FT	STRAND	533	534
FT	TURN	537	538
FT	STRAND	540	542
FT	HELIX	543	545
FT	TURN	546	546
FT	STRAND	550	550
FT	TURN	552	553
FT	STRAND	555	557
FT	HELIX	561	565
FT	TURN	566	567
FT	HELIX	569	572
FT	TURN	573	573
SO	SEQUENCE	602 AA;	63953 MW; 982EF3245D87C43E CRC64;

Query Match 11.8%; Score 91.5; DB 1; Length 602;
Best Local Similarity 41.7%; Pred. No. 1.6;
Matches 25; Conservative 2; Mismatches 12; Indels 21; Gaps 2;

QY 81 YDQVTRVTHMAHAF-----RNNATTIDOMAKNSDITY-----GOYG 119
Db 254 YDQVTRVTHMAHAFSGPFEDARIVANVPVRGKNPDVPVINSSTAVAKAREQYG 313

RESULT 6
CSGB_ECOLI STANDARD; PRT; 151 AA.
ID CSGB_ECOLI
AC P39828;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Minor curlin subunit precursor.
OS CSGB OR B1041 OR Z1675 OR ECS1419.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MC4100;
RX MEDLINE=96414468; PubMed=8817489;
RA Hammar M., Arngqvist A., Bian Z., Olsen A., Normark S.;
RT "Expression of two csg operons is required for production of
RT fibronectin- and congo red-binding curli polymers in Escherichia coli
RT K-12.";
RL Mol. Microbiol. 18:661-670(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";

Science 277:1453-1474(1997).

[3]

RP SEQUENCE FROM N.A.

RC STRAIN=K12;

RX MEDLINE=97061202; PubMed=8905232;

RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A., Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kishimoto K., Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K., Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N., Samped G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y., Yano M., Horuchi T.;

RT "A 718-kb DNA sequence of the *Escherichia coli* K-12 genome corresponding to the 12.7-28.0 min region on the linkage map.;"

RL DNA Res. 3:137-155(1996).

[4]

RP SEQUENCE FROM N.A.

RC STRAIN=O157:H7 / EDL933 / ATCC 700927;

RX MEDLINE=21074935; PubMed=11206551;

RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glaesner J.D., Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A., Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L., Grobbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoudis K., Apodaca U., Anantharaman T.S., Iain J., Yen G., Schwartz D.C., Welch R.A., Blattner F.R.;

RT "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7.;"

RL Nature 409:529-533(2001).

[5]

RP SEQUENCE FROM N.A.

RC STRAIN=O157:H7 / RIMD 0509952;

RX MEDLINE=21156231; PubMed=11258796;

RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K., Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T., Iida T., Takami H., Honda T., Sasakiwa C., Ogasawara N., Yasunaga T., Kubara S., Shiba T., Hattori M., Shinagawa H.;

RT "Complete genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12.;"

RL DNA Res. 8:11-22(2001).

[6]

RP SEQUENCE OF 1-21 FROM N.A.

RC STRAIN=K12;

RX MEDLINE=95157246; PubMed=7854117;

RA Arrqvist A., Olsen A., Normark S.;

RT "Sigma S-dependent growth-phase induction of the *csgB* promoter in *Escherichia coli* can be achieved in vivo by sigma 70 in the absence of the nucleoid-associated protein H-NS.;"

RL Mol. Microbiol. 13:1021-1032(1994).

CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO FIBROBLASTS. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF CURLIN MONOMERS.

CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC

DR EMBL; X90754; CA62281.1; -

DR EMBL; AE000205; AAC74125.1; -

DR EMBL; D90741; BAA35831.1; -

DR EMBL; AE005315; AAG55767.1; -

DR EMBL; AF002554; BAB34842.1; -

DR PIR; C90806; C90806.

DR PIR; G85665; G85665.

DR PIR; S70787; S70787.

DR EcoGene; EG12621; csgB

KM FimD; Signal; Complete proteome.

FT CHAIN 1 21 POTENTIAL.

FT CHAIN 22 151 MINOR CURLIN SUBUNIT.

SQ SEQUENCE 151 AA; 15882 MW; B18D26B964014B8 CRC64;

Query Match 11.5%; Score 89.5; DB 1; Length 151;

Best Local Similarity 28.1%; Pred. No. 0.5;

Matches 32; Conservative 15; Mismatches 62; Indels 5; Gaps 2;

QY 38 SSGPSTLSIYOGGANALALQSDARKSETTITOSGYGNGADVDQVTRVVTHEMAAF 97

DB 21 AAGYDLANEYFNF-----AVNELSKSFNQALIGAGTNSQALQGGSKLLAVVAQSGS 76

QY 98 RNNATIDQNNAKNSDITVGYGNNALVNOTASDSSVWVROVGFGNNTANQY 151

DB 77 SRRAKIDQTDGYNL-AVIDQAGSANDASISQAYGNTAMTIQSGGNKANTIOY 129

RESULT 7

ID ICEA_PANAN STANDARD; PRT; 1322 AA.

AC P20469;

DT 01-FEB-1991 (Rel. 17, Created)

DT 01-FEB-1991 (Rel. 17, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Ice nucleation protein inaA.

GN INAA.

OS *Pantoea ananas* (Erwinia uredovora).

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Pantoea.

OX NCBI_Taxid=553;

RN (1)

RP SEQUENCE FROM N.A.

RX MEDLINE=90092494; PubMed=2599095;

RA Abe K., Watabe S., Emori Y., Watanabe M., Arai S.;

RT "An ice nucleation active gene of *Erwinia ananas*. Sequence similarity to those of *Pseudomonas* species and regions required for ice nucleation activity.;"

RL FEBS Lett. 258:297-300(1989).

CC -1- FUNCTION: Ice nucleation proteins enable bacteria to nucleate crystallization in supercooled water.

CC -1- SUBCELLULAR LOCATION: Outer membrane (By similarity).

CC -1- DOMAIN: CONTAINS MANY IMPERFECT REPEATS OF THE CONSENSUS OCTAPEPTIDE A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.

CC -1- SIMILARITY: Belongs to the bacterial ice nucleation protein family.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC

DR EMBL; X17316; CAA55194.1; -

DR PIR; S07053; S07053.

DR HSRP; P06620; INAA.

DR InterPro; IPR000258; Ice_nucleatn.

DR Pfam; PF00818; Ice_nucleation; 69.

DR PRINTS; PR00327; ICENUCLEATN.

DR PROSITE; PS00314; ICE_NUCLEATION; 49.

KW Ice nucleation; Repeat; Outer membrane.

FT DOMAIN 162 1281 OCTAPEPTIDE PERIODICITY.

SQ SEQUENCE 1322 AA; 131094 MW; 89B0BE24AAB37039 CRC64;

Query Match 11.5%; Score 89.5; DB 1; Length 1322;

Best Local Similarity 26.8%; Pred. No. 5.9;

Matches 34; Conservative 18; Mismatches 26; Indels 49; Gaps 7;

QY 34 GGNSSGPSTLSIYOGGANALALQSDARKSETTITOSGYGNGADVDQVTRVVTHEM 93

DB 933 GSTSTAGPSSSL-INGYSTQTA-----GNVSLTAGYS----- 966

QY 94 AAHAFNNATIDQNNAKNSDITVGYG-----GNNALV-----NOTASDSSVWVROVGF 143

Db 967 -----TGTGENSEDLITG-YGSTSTAGTSSLIAGYSTGTATSPKSTLM-AGVG 1013

QY 144 NNATANO 150

Db 1014 SSGTARE 1020

RESULT 8

ICEN_XANCT STANDARD; PRT; 1567 AA.

ID ICEN_XANCT P18127;

AC 01-NOV-1990 (Rel. 16, Created)

DT 01-NOV-1990 (Rel. 16, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Ice nucleation protein.

GN INAX.

OS Xanthomonas campestris (pv. translucens).

OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;

OC Xanthomonadaceae; Xanthomonas.

OX NCBI_TaxID=343;

RN (1)

RP SEQUENCE FROM N.A.

RC STRAIN=X565;

RX MEDLINE=91080859; PubMed=2259339;

RA Zhao J., Orser C.S.;

RT "Conserved repetition in the ice nucleation gene inax from

RT Xanthomonas campestris pv. translucens.";

RL Mol. Genet. 223:163-166(1990).

CC -1- FUNCTION: Ice nucleation proteins enable bacteria to nucleate

CC crystallization in supercooled water.

CC -1- SUBCELLULAR LOCATION: Outer membrane (By similarity).

CC -1- DOMAIN: CONTAINS 153 IMPERFECT REPEATS OF THE CONSENSUS

CC OCTAPEPTIDE A-G-Y-G-S-T-L-T; FURTHER ON A 16-RESIDUE AND A

CC REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.

CC -1- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE

CC NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.

CC -1- SIMILARITY: Belongs to the bacterial ice nucleation protein

CC family.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL: X52970; CAA37140.1; -.

CC HSSP: P06620; 11NA.

CC InterPro: IPR000258; Ice_nucleatn.

CC Pfam: PF00818; Ice_nucleatn; 81.

CC PRINTS: PR00327; ICENUCLEATN.

CC PROSITE: PS00314; ICE_NUCLEATION; 57.

CC Ice_nucleation; Repeat; Outer membrane.

CC KW SEQUENCE 1567 AA; 152548 MW; C8B451D959ECAD63 CRC64;

CC SQ

Query Match 11.5%; Score 89.5; DB 1; Length 1567;

Best local Similarity 26.4%; Pred. No. 7.2; Indels 49; Gaps 12;

Matches 47; Conservative 23; Mismatches 59; Indels 49; Gaps 12;

QY 14 VSGSALAG-----VFPOMG--GNNM-----GGNSGSPDSTLSIYOGSANAAL 59

Db 205 VVGSTLTGADQSRVLVAGYGTETAGDHSPLIAGYGTGAGSDSI-LAGYGTGTAAAG 263

QY 60 Q-----SPARSETTITGSGY--NGAD-----YDQVTRVVTHEMAAFRNAT 102

Db 264 STLTAGYGTQTAAQESRLTSGYSTATSGSDSAVTSGYSTGTAGSSESLTAGYSTGT 323

QY 103 IDQMANKSDITVGOYG-----GNNAALV-----NOTASDSVWVRQVGFNNATNO 150

Db 324 A-----RKGSDITAG-YGSTGTAGSDSALLIAGYSTGTAGSESLT--AGVSTGTAAK 374

RESULT 9

ICEN_PANAN STANDARD; PRT; 1034 AA.

ID ICEN_PANAN Q47879;

AC 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Ice nucleation protein Inau.

GN INAU.

OS Pantoea ananas (Erwinia uredovora).

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Pantoea.

OX NCBI_TaxID=553;

RN (1)

RP SEQUENCE FROM N.A.

RC STRAIN=KUN-3;

RX MEDLINE=94264407; PubMed=7764866;

RA Michigami Y., Watabe S., Abe K., Odata H., Arai S.;

RT "Cloning and sequencing of an ice nucleation active gene of Erwinia

RT uredovora.";

RL Biosci. Biotechnol. Biochem. 58:762-764(1994).

CC -1- FUNCTION: Ice nucleation proteins enable bacteria to nucleate

CC crystallization in supercooled water.

CC -1- SUBCELLULAR LOCATION: Outer membrane.

CC -1- DOMAIN: CONTAINS IMPERFECT REPEATS OF A CONSENSUS OCTAPEPTIDE

CC A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A REGIONAL 48-RESIDUE

CC PERIODICITY IS SUPERIMPOSED.

CC -1- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE

CC NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.

CC -1- SIMILARITY: Belongs to the bacterial ice nucleation protein

CC family.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL: D14992; BAA03636.1; -.

CC FIR: JC2143; JC2143.

CC HSSP: P06620; 11NA.

CC InterPro: IPR000258; Ice_nucleatn.

CC Pfam: PF00818; Ice_nucleation; 51.

CC PRINTS: PR00327; ICENUCLEATN.

CC PROSITE: PS00314; ICE_NUCLEATION; 34.

CC Ice_nucleation; Repeat; Outer membrane.

CC KW DOMAIN 162 993 OCTAPEPTIDE PERIODICITY.

CC FT SEQUENCE 1034 AA; 103378 MW; FA222523D333EADD CRC64;

CC SQ

Query Match 11.3%; Score 87.5; DB 1; Length 1034;

Best local Similarity 25.2%; Pred. No. 6.6; Indels 63; Gaps 9;

Matches 41; Conservative 18; Mismatches 41; Indels 63; Gaps 9;

QY 12 IVVSGSALAGVYPPW--GGGNNHNGGNS-----GPDSTLSIYOGSANAAL 57

Db 161 INTVSTLTGTHQSLIAGYSTETAGDSSTLIAGYSTGTAGSDSTL-VAGYGTQTAA 218

QY 58 ALQSDARSETTITGSGYGNAGDYDQVLTTRVVTHEMAAFRNATITDQMANKSDITVQ 117

Db 219 -----GEBSQMAAGYS-----TGTGKSGSDITAG- 243

QY 118 VG-----GNNAALV-----NOTASDSVWVRQVGFNNATNO 150

Db 244 YSTGTAGDSSLIAGYSTGTAGSDSLLT--AGVSTGTAAK 284

RESULT 10

ICEN_ERWHE STANDARD; PRT; 1258 AA.

ID ICEN_ERWHE P16239;

DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Ice nucleation protein.
 OS Erwinia herbicola.
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 CC Enterobacteriaceae; Pantoea.
 OX NCBI_TaxID=549;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=M1;
 RX MEDLINE=90152370; PubMed=2515997;
 RA Warren G.J., Corotto L.V.;
 RT "The consensus sequence of ice nucleation proteins from *Erwinia herbicola*, *Pseudomonas fluorescens* and *Pseudomonas syringae*."
 RL Gene 85:239-242(1989)
 CC -1- FUNCTION: Ice nucleation proteins enable bacteria to nucleate
 CC crystallization in supercooled water.
 CC -1- SUBCELLULAR LOCATION: Outer membrane.
 CC -1- DOMAIN: CONTAINS 126 IMPERFECT REPEATS OF A CONSENSUS OCTAPEPTIDE
 CC A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A REGIONAL 48-RESIDUE
 CC PERIODICITY IS SUPERIMPOSED
 CC -1- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE
 CC NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.
 CC -1- SIMILARITY: Belongs to the bacterial ice nucleation protein
 CC family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M26382; AAA24823.1; -
 DR PIR: J00188; J00188.
 DR HSSP: F06620; INA.
 DR InterPro: IPR000258; Ice nucleatn.
 DR Pfam: PF00818; Ice nucleation; 65.
 DR PRINTS: PRO0327; ICENUCLEATN.
 DR PROSITE: PS00314; ICE NUCLEATION; 45.
 KW Ice nucleation; Repeat; Outer membrane.
 KM DOMAIN 162 1217 OCTAPEPTIDE PERIODICITY.
 FT SEQUENCE 1258 AA; 125084 MW; 59088A130077EBD4 CRC64;
 SQ
 Query Match 11.3%; Score 87.5; DB 1; Length 1258;
 Best Local Similarity 25.2%; Pred. No. 8.2;
 Matches 41; Conservative 18; Mismatches 41; Indels 63; Gaps 9;
 QY 12 IVVSGSALAGVPPW--GGGNNHNGGNS-----GPDSTLSIYOGSANAAL 57
 DB 161 IATVSTLSGTHSQSLAGVSTETAGDSTLLAGYSTGTAGADSTL-VAGVSGTQA- 218
 QY 58 ALQSDARKSETTITQSGNGADVDLVTRVYTHEMAAFRNATIDQNAKSDITVQ 117
 DB 219 -----GESSQMAVGS-----TQTMKSGDITLAG- 243
 QY 118 YG-----GNNAAV-----NOTASDSSVMVRQVFGNNAATQ 150
 DB 244 YGSGTAGDSSSLTAGVSGTQTAGESSLT--AGVSGTQTAK 264
 RESULT 11
 ID CSGB_SALTI STANDARD; PRT; 151 AA.
 AC Q827M3;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Minor curlin subunit precursor.
 CSGB OR STY1180 OR T1777.

OS Salmonella typhi.
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 CC Enterobacteriaceae; Salmonella.
 OX NCBI_TaxID=601;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CT18;
 RX MEDLINE=21534947; PubMed=11677608;
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
 RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
 RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
 RA Cronin A., Davis P., Davies R.M., Dowd L., Farrar J.,
 RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holtroyd S., Jagels K.,
 RA Kirog A., Larsen T.S., Leather S., Moulie S., O'Garra P., Parry C.,
 RA Quail M.A., Rutherford K., Simmonds M., Skellon J., Stevens K.,
 RA Whitehead S., Barrrell B.G.;
 RT "Complete genome sequence of a multiple drug resistant *Salmonella*
 RT enterica serovar Typhi CT18."
 RL Nature 413:848-852(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Ty2 / ATCC 700931;
 RX MEDLINE=22531367; PubMed=12644504;
 RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
 RA Burland V., Kodyanmi V., Schwartz D.C., Blatter F.R.;
 RT "Comparative genomics of *Salmonella enterica* serovar Typhi strains Ty2
 RT and CT18."
 RL J. Bacteriol. 185:2330-2337(2003).
 CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
 CC CURLIN MONOMERS.
 CC -----
 CC -1- SIMILARITY: BELONGS TO THE CSGB/CSGB FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AL627269; CAD08267.1; -
 DR EMBL: AE016840; AA069400.1; -
 KW Fimbria; Signal; Complete proteome.
 FT SIGNAL 1 21
 FT CHAIN 1 151
 FT SEQUENCE 151 AA; 16254 MW; 161C54326E573495 CRC64;
 SQ
 Query Match 11.2%; Score 86.5; DB 1; Length 151;
 Best Local Similarity 26.2%; Pred. No. 0.89;
 Matches 34; Conservative 17; Mismatches 42; Indels 37; Gaps 4;
 QY 10 AAIIVSSALAGVPPWGGGNNHNGGNSGPDSTLSIYOGSANAALALQSDARKSETT 69
 DB 58 ARVQEGSKLISVLSQ--EGENNRKAVDQGNVFAVIEQGNAN-----DAS 103
 QY 70 ITQSGYNGADVDLVTRVYTHEMAAFRNATIDQNAKSDITVQYQYGNNAALVNOT 129
 DB 104 ISQAYGNSA-----AIIQKSGNKNANT--QYGTQTAIVVQK 140
 QY 130 ASDSSVMVRQ 139
 DB 141 QSHMAIRVQ 150
 RESULT 12
 ID GP63_LEIDO STANDARD; PRT; 590 AA.
 AC P23223;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)

FT CARBOHYD 433 433 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 445 445 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 466 466 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 501 501 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 646 AA; 63054 MW; FE48DDC78C10B0A CRC64;

Query Match 10.8%; Score 84.5; DB 1; Length 646;
 Best Local Similarity 36.5%; Pred. No. 6.9;
 Matches 23; Conservative 7; Mismatches 12; Indels 21; Gaps 2;

QY 81 YDOLVTVTHMAHAFRRNAT-----IDONNAKNSDTV-----GQYG 119
 Db 256 YDOLVTVTHMAHAFRRNAT-----IDONNAKNSDTV-----GQYG 119
 QY 120 GNN 122
 Db 316 CNS 318

RESULT 14
 GP63_LEICH STANDARD; PRT; 599 AA.
 AC P15706;
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Leishmania gp63 precursor (EC 3.4.24.36) (Cell surface protease)
 DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface endopeptidase).
 GN GP63.
 OS Leishmania chagasi.
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 OX NCBI_TaxID=44271;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90205976; PubMed=2320059;
 RA Miller R.A., Reed S.G., Parsons M.;
 RT "Leishmania gp63 molecule implicated in cellular adhesion lacks an Arg-Gly-Asp sequence."
 RL J. Biol. Chem. 267:1888-1895(1992).
 RN (2)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92112918; PubMed=1370484;
 RA Ramamoorthy R., Donelson J.E., Paetz K.E., Maybodi M., Roberts S.C., Wilson M.E.;
 RT "Three distinct RNAs for the surface protease gp63 are differentially expressed during development of Leishmania donovani chagasi promastigotes to an infectious form."
 RL J. Biol. Chem. 267:1888-1895(1992).
 CC -1- FUNCTION: Has an integral role during the infection of macrophages in the mammalian host.
 CC -1- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and P1 and basic residues at P2 and P3'. A model nonapeptide is cleaved at -Ala-Tyr-|-Leu-Lys-Lys-.
 CC -1- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
 CC -1- SIMILARITY: Belongs to peptidase family M8.

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

 DR EMBL; M80672; AAA29238.1; -
 DR EMBL; M28527; AAA29235.1; -
 DR PIR; A44951; A44951.
 DR HSSP; P08148; 1LM1.
 DR MEROPS; M08.001.
 DR InterPro; IPR006025; Pept_Mzn_BS.
 DR InterPro; IPR001577; Peptidase_M8.

DR Pfam; PF01457; Peptidase_M8; 1.
 DR PRINTS; PR00782; LSHMANOLYSIN.
 DR PROSITE; PS00142; ZINC_PROTEASE; 1.
 KW Hydrolase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;
 KW Zymogen; Signal; Cell adhesion; GPI-anchor; Lipoprotein.
 FT SIGNAL 1 39
 FT PROPEP 40 97
 FT CHAIN 98 574
 FT PROPEP 575 599
 FT METAL 261 261
 FT ACT SITE 262 262
 FT METAL 265 265
 FT METAL 331 331
 FT DISULFID 122 139
 FT DISULFID 188 227
 FT DISULFID 311 383
 FT DISULFID 390 452
 FT DISULFID 403 422
 FT DISULFID 412 486
 FT DISULFID 463 507
 FT DISULFID 512 562
 FT DISULFID 532 555
 FT CARBOHYD 297 394
 FT CARBOHYD 394 574
 FT LIPID 574 574

Query Match 10.8%; Score 84; DB 1; Length 599;
 Best Local Similarity 37.5%; Pred. No. 7;
 Matches 24; Conservative 5; Mismatches 29; Indels 6; Gaps 1;

QY 39 SGPDSLTSLIYQGSNMAA-----LQSDAKSETTTQSGYGNAGDYDOLTVRVTH 92
 Db 203 SNTDFVWYASVPSEGVLAWATTCQVPSDHPAGVINIPANIASRYDOLTVRVTH 262
 QY 93 MAHA 96
 Db 263 MAHA 266

RESULT 15
 HLVA_SERMA STANDARD; PRT; 1608 AA.
 AC P15320;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update).
 DT 01-NOV-1990 (Rel. 16, Last annotation update)
 DE Hemolysin precursor.
 GN SHLA.
 OS Serratia marcescens.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Serratia.
 OX NCBI_TaxID=615;
 RN (1)
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 31-40.
 RX STRAIN=SN8;
 RX MEDLINE=88257037; PubMed=3290200;
 RA Poole K., Schiebel E., Braun V.;
 RT "Molecular characterization of the hemolysin determinant of Serratia marcescens."
 RL J. Bacteriol. 170:3177-3188(1988).
 CC -1- FUNCTION: Bacterial hemolysins are exotoxins that attack blood cell membranes and cause cell rupture by mechanisms not clearly defined.
 CC -1- FUNCTION: CELL-BOUND HEMOLYSIN, WHICH RELEASES HEME-IRON FROM ERYTHROCYTES BY INTERACTION WITH THE ERYTHROCYTE MEMBRANE. SHLA REQUIRES SHIB FUNCTION.
 CC -1- SUBCELLULAR LOCATION: Outer membrane.
 CC -1- SIMILARITY: TO P.MIRABILIS HEMOLYSIN (HPMA).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; M22618; AAA50323.1; -.
 DR PIR; A28182; A28182.
 DR InterPro; IPR008638; Haemag_act.
 DR Pfam; PF05860; Haemag_act; 1.
 KM Hemolysate; Toxin; Outer membrane; Signal.
 FT SIGNAL 1 30
 FT CHAIN 31 1608 HEMOLYSIN.
 SQ SEQUENCE 1608 AA; 165078 MW; D669B476FE7DAD51 CRC64;

Query Match 10.8%; Score 84; DB 1; Length 1608;
 Best Local Similarity 27.0%; Pred. No. 22;
 Matches 33; Conservative 16; Mismatches 53; Indels 20; Gaps 4;

OY 29 GGNHNGGSSGPDSTLSIYQGSANAALALQSDARKSETTTTQSGYGADYD-QLVTR 87
 DB 842 GAQHSVCG-----AYQENAGVDHLAADTAATTTTQTDVGVIQANVDYSAVTR 891
 OY 88 VYTHEMAHAFRNAT-----IDQNAKNSDITVQYCGNNAALVNOFASDSSVMROYGF 142
 DB 892 PVERAVGKAAKLDATGVINDIGGIGAPNVGLDIGAQGSSS-----EKRSSSSQAIVSSVQA 947
 OY 143 GN 144
 DB 948 GS 949

Search completed: March 11, 2004, 18:34:59
 Job time : 6.3 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:23:54 ; Search time 30.5 Seconds
(without alignments)
1562.074 Million cell updates/sec

Title: US-09-543-407-28
Perfect score: 775
Sequence: 1 MKLLKVAAPAIIVSGSALA.....DSVMVRQVFGNNATANQY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

```
Searched:      1017041 begs, 315518202 residues
Total number of hits satisfying chosen parameters: 10170411
```

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

```
Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

```
Database :
SPRMBL_25.*
1: sp archaea.*
2: sp bacteria.*
3: sp fungi.*
4: sp human.*
5: sp invertebrate.*
6: sp mammal.*
7: sp mhc.*
8: sp organelle.*
9: sp phage.*
10: sp plant.*
11: sp rodent.*
12: sp virus.*
13: sp vertebrate.*
14: sp unclassified.*
15: sp virus.*
16: sp_bacteriap.*
17: sp_archaeap.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysts of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	674	87.0	152	2	O33802	O33802 salmonella
2	566.5	73.1	150	2	Q7X433	Q7X243 citrobacter
3	530	68.4	149	2	Q7X240	Q7X240 citrobacter
4	484.5	62.5	152	16	Q8CW63	Q8CW63 escherichia
5	411.5	53.1	150	2	Q7X237	Q7X237 enterobacte
6	295	38.1	76	2	Q54069	Q54069 salmonella
7	122	15.7	29	2	Q9S3J5	Q9S3J5 escherichia
8	112	14.5	130	16	Q8SJ14	Q8SJ14 bradyrhizob
9	111.5	14.4	3501	16	Q8Y106	Q8Y106 ralstonia s
10	111.5	14.4	3552	16	Q8XSD6	Q8XSD6 ralstonia s
11	110	14.2	139	16	Q8R1D3	Q8R1D3 shewanella
12	108	13.9	1432	16	Q8EFU3	Q8EFU3 shewanella
13	107	13.8	2174	16	Q9ZU08	Q9ZU08 rhizobium m
14	106.5	13.7	502	16	Q8R1H4	Q8R1H4 shewanella
15	99.5	12.8	171	16	Q8SJ13	Q8SJ13 bradyrhizob
16	99	12.8	74	16	Q7U5X6	Q7U5X6 synechococc

17	98	12.6	179	2	Q3801	03801 salmonella
18	97.5	12.6	154	5	Q89J75	Q89J15 bradyrhizob
19	97	12.5	1748	5	Q94821	Q94821 tetrahymena
20	95	12.3	368	16	Q8EW6	Q8EW6 mycoplasma
21	98.5	12.3	400	10	Q8L4R8	Q8L4R8 oryza sativ
22	95	12.3	1410	16	Q8CMJ0	Q8CMJ0 shewanella
23	94.5	12.2	140	16	Q8CW64	Q8CW64 escherichia
24	94	12.1	362	16	Q8BV84	Q8BV84 mycoplasma
25	94	12.1	645	16	Q7U1C5	Q7U1C5 mycobacteri
26	94	12.1	646	16	Q53B18	Q53B18 mycobacteri
27	93.5	12.1	644	5	Q8NMV9	Q8NM9 leishmania
28	93	12.0	157	16	Q8BHG0	Q8BHG0 pseudomonas
29	92.5	11.9	348	13	Q93397	Q93397 cyprinus ca
30	92.5	11.9	644	5	Q43994	Q43994 leishmania
31	92.5	11.9	1765	5	Q7V8S5	Q7V8S5 prochloroco
32	92	11.9	598	5	Q25275	Q25275 leishmania
33	91	11.7	152	2	Q7X241	Q7X241 clyrobacter
34	91	11.7	53	2	Q9SAF2	Q9SAF2 arabidopsis
35	91	11.7	614	16	Q7TVR8	Q7TVR8 mycobacteri
36	91	11.7	615	16	P95249	P95249 mycobacteri
37	91	11.7	1408	16	Q8EB83	Q8EB83 shewanella
38	90.5	11.7	1266	5	Q841Y5	Q841Y5 campylobact
39	89.5	11.5	151	2	Q7X238	Q7X238 enterobacte
40	89.5	11.5	151	16	Q7UCZ1	Q7UCZ1 shigella fl
41	89.5	11.5	153	16	Q89JUV	Q89J16 bradyrhizob
42	89.5	11.5	160	16	Q83RU7	Q83RU7 shigella fl
43	89.5	11.5	1333	16	Q8PD38	Q8PD38 xanthomonas
44	89.5	11.5	2411	16	Q8PF72	Q8PF72 xanthomonas
45	89	11.5	354	16	P95248	P95248 mycobacteri

ALIGNMENTS

Query Match	Best Local Similarity	Matches 134;	Conservative 2;	Mismatches 15;	Indels 0;	Gaps 0;
Qy	87.0%; Score 674; DB 2; Length 152;					
Ds	1 MKLLKVAFAFAIVVSGSALAGVVPWGGGNGNHNGGSSGPDSTLSTIYOGSANAALALQ 60					
Ds	1 MKLLKVAFAFAIVVSGSALAGVVPWGGGNGNHNGGSSGPDSTLSTIYOGSANAALALQ 60					
Qy	61 SDARKSETTTTQSGYGGAGADYDQLVTRVVTHEMAHARRNATTIDQWNAKXSDITVGGYGG 120					
Ds	61 SDARKSETTTTQSGYGGAGADYDQLVTRVVTHEMAHARRNATTIDQWNAKXSDITVGGYGG 120					
Qy	121 NNALVNVGTASDSSVMVRQVGFGNATATANCY 151					

```

Db      121 NNAALVNOTASDSSVWVRQVGFNNATANQY 151
      |||
RESULT 2
Q7X243 PRELIMINARY; PRT; 150 AA.
AC Q7X243;
DT 01-OCT-2003 (TREMBLrel. 25, Created)
DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)
DE 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Curlin-csgA protein.
GN CSGA.
OS Citrobacter sp., Fec2.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Citrobacter.
OX NCBI_TaxID=213763;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fec2;
RA Zogaj X., Bokranz W., Nimtz M., Romling U.;
RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RL Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
EMBL: AJ515700; CAD56672.1; -.
DR EMBL: AJ515700; CAD56672.1; -.
SQ SEQUENCE 150 AA; 15016 MW; 1D7141B8D6973DC6 CRC64;

Query Match      73.1%; Score 566.5; DB 2; Length 150;
Best Local Similarity 76.8%; Pred. No. 6.5e-39;
Matches 116; Conservative 10; Mismatches 24; Indels 1; Gaps 1;

Qy      1 MLLKVAAPAAIVVSGSALAGVPPWGGGNNHGGSSGPDSTLSIYOGSANAALALQ 60
Db      1 MLLQVAAPAAIVVSGSALAGVPPWGGGGG-GGGSSSGSESTLSIYOGSVNNAALALQ 59
Qy      61 SDARKSETTTTQSGYGNGADVDQLVTRVVTHEMAAFNNATTIDQWNAKNSDITVQYGG 120
Db      60 SDARKSDTTTHQNGGNGADVGQSGSDNSTIDLTQNGFGNNATTIDQWNGKNSDITVQYGG 119

Qy      121 NNAALVNOTASDSSVWVRQVGFNNATANQY 151
Db      120 HNAALVNOTASDSSVWVRQVGFNNATANQY 150

RESULT 3
Q7X240 PRELIMINARY; PRT; 149 AA.
AC Q7X240;
DT 01-OCT-2003 (TREMBLrel. 25, Created)
DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)
DE 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Curlin-csgA protein.
GN CSGA.
OS Citrobacter freundii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Citrobacter.
OX NCBI_TaxID=3546;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fec4;
RA Zogaj X., Bokranz W., Nimtz M., Romling U.;
RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RL Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
EMBL: AJ515701; CAD56675.1; -.
DR EMBL: AJ515701; CAD56675.1; -.
SQ SEQUENCE 149 AA; 15260 MW; 946DD52017F648FD CRC64;

Query Match      68.4%; Score 530; DB 2; Length 149;
Best Local Similarity 71.5%; Pred. No. 6.3e-36;
Matches 108; Conservative 16; Mismatches 25; Indels 2; Gaps 1;

Qy      1 MLLKVAAPAAIVVSGSALAGVPPWGGGNNHGGSSGPDSTLSIYOGSANAALALQ 60

```

```

Db      1 MLLKVAAPAAIVVSGSALAGVPPW--GGNHGGSSNYGPPDSLSIYOGSNNANALQ 58
Qy      61 SDARKSETTTTQSGYGNGADVDQLVTRVVTHEMAAFNNATTIDQWNAKNSDITVQYGG 120
Db      59 SDARKSDTTTHQNGGNGAGVAGQGADSDTSLKQTFQNSATIDQWNAKADISVTFQFG 118

Qy      121 NNAALVNOTASDSSVWVRQVGFNNATANQY 151
Db      119 RQNALVNOTASDSSVWVRQVGFNNATANQY 149

RESULT 4
Q8CW63 PRELIMINARY; PRT; 152 AA.
AC Q8CW63;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DE Major curlin subunit precursor.
GN CSGA OR C1306.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O6.HI / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).
DR EMBL: AE016759; AA079779.1; -.
KW Complete proteome.
SQ SEQUENCE 152 AA; 15064 MW; 3BA57F34C1240E83 CRC64;

Query Match      62.5%; Score 484.5; DB 16; Length 152;
Best Local Similarity 66.4%; Pred. No. 3.4e-32;
Matches 101; Conservative 15; Mismatches 35; Indels 1; Gaps 1;

Qy      1 MLLKVAAPAAIVVSGSALAGVPPW--GGGNNHGGSSGPDSTLSIYOGSANAALALQ 59
Db      1 MLLKVAAPAAIVVSGSALAGVPPYGGGGNNHGGSSGPNSELTNIYQGGNSALALQ 60
Qy      60 QSDARKSETTTTQSGYGNGADVDQLVTRVVTHEMAAFNNATTIDQWNAKNSDITVQYGG 119
Db      61 QADARNSDTTTHQGGGNGADVGQSGSDSDSIDLTQRFGNAGATLDQWNGKSDITVQYFG 120

Qy      120 GNAALVNOTASDSSVWVRQVGFNNATANQY 151
Db      121 GNGAALVNOTASDSSVWVRQVGFNNATANQY 152

RESULT 5
Q7X237 PRELIMINARY; PRT; 150 AA.
AC Q7X237;
DT 01-OCT-2003 (TREMBLrel. 25, Created)
DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)
DE 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Curlin-csgA protein.
GN CSGA.
OS Enterobacter sakazakii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Enterobacter.
OX NCBI_TaxID=28141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fec39;
RA Zogaj X., Bokranz W., Nimtz M., Romling U.;

```



```

DT 01-OCT-2003 (Tremblrel. 25, last annotation update)
DE Probable hemagglutinin-related protein.
GN RSC0887 OR RS06116
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
ON NCBI_TaxID=305;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RA MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Chandler M., Choise N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
RA Signier P., Thebaud P., Whalen M., Wincker P., Levy M.,
RA Weisenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502 (2002).
DR EMBL; AL646061; CAD14589.1; -
DR GO; GO:0004519; F:endonuclease activity; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001604; Endonuclease.
DR InterPro; IPR008619; F11_haemag.
DR Pfam; PF05594; F11_haemag; 20.
DR Pfam; PF05860; Haemag_act; 1.
DR PROSITE; PS01070; NUCLEASE_NON_SPEC; 1.
DR Complete proteome.
SQ SEQUENCE 3501 AA; 348421 MW; 290B41C99018A107 CRC64;

Query Match 14.4%; Score 111.5; DB 16; Length 3501;
Best Local Similarity 29.3%; Pred. No. 5.2;
Matches 41; Conservative 20; Mismatches 52; Indels 27; Gaps 5;

QY 15 SSSALAGVPPWGGGNNHGG-NSSGPDSTLSTIYOGSANA----- 56
DB 2431 SSSHFSTAGPSWCDLERNVGGGPNSSG---VGLAPYGSASADNAAGNSRQNASVYIG 2486
QY 57 LALQSDARKSETTITGSGYGADYDQVTVRVTHMHAFFNNATTIDQWNAKNSDITVG 116
DB 2487 KSVQVQARTGDTIVSGSGISALSDVDLAKQKQVIVA---GNDTSSR-HEHSDRTIG 2541
QY 117 QYGGNNAALVNOTASDSSVM 136
DB 2542 DLGGNGSGTVGVRASSTL 2561

RESULT 10
Q8XSD6 PRELIMINARY; PRT; 3552 AA.
AC Q8XSD6;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 25, last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, last annotation update)
DE Probable hemagglutinin-related protein.
GN RS05040 OR RS06117.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Plasmid megaplasmid.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
ON NCBI_TaxID=305;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RA MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Chandler M., Choise N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
RA Signier P., Thebaud P., Whalen M., Wincker P., Levy M.,
RA Weisenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";

```

```

RL Nature 415:497-502 (2002).
DR EMBL; AL646079; CAD17691.1; -
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
DR GO; GO:0004519; F:endonuclease activity; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001604; Endonuclease.
DR InterPro; IPR008619; F11_haemag.
DR Pfam; PF05594; F11_haemag; 20.
DR Pfam; PF05860; Haemag_act; 1.
DR PROSITE; PS01070; NUCLEASE_NON_SPEC; 1.
DR Plasmid; Complete proteome.
SQ SEQUENCE 3552 AA; 352934 MW; C5432AAB2CCCF59C CRC64;

Query Match 14.4%; Score 111.5; DB 16; Length 3552;
Best Local Similarity 29.3%; Pred. No. 5.3;
Matches 41; Conservative 20; Mismatches 52; Indels 27; Gaps 5;

QY 15 SSSALAGVPPWGGGNNHGG-NSSGPDSTLSTIYOGSANA----- 56
DB 2430 SSSHFSTAGPSWCDLERNVGGGPNSSG---VGLAPYGSASADNAAGNSRQNASVYIG 2485
QY 57 LALQSDARKSETTITGSGYGADYDQVTVRVTHMHAFFNNATTIDQWNAKNSDITVG 116
DB 2486 KSVQVQARTGDTIVSGSGISALSDVDLAKQKQVIVA---GNDTSSR-HEHSDRTIG 2540
QY 117 QYGGNNAALVNOTASDSSVM 136
DB 2541 DLGGNGSGTVGVRASSTL 2560

RESULT 11
Q8EHI3 PRELIMINARY; PRT; 139 AA.
AC Q8EHI3;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, last annotation update)
DE Minor curlin subunit CsgB, putative.
GN SC00866.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.
ON NCBI_TaxID=70863;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=NR-1;
RA MEDLINE=22297686; PubMed=12368813;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson R.A.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Dougherty S.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Umeyam L.A., White O., Wolf A.M.,
RA Vamathevan J., Weidman J., Imprim M., Lee K., Berry K., Lee C.,
RA Mueller H., Kouri H., Gill J., Utecherback T.R., McDonald L.A.,
RA Feldblyum T.V., Smith H.O., Venter J.C., Neilson K.H., Fraser C.M.;
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
Shewanella oneidensis.";
RL Nat. Biotechnol. 20:1118-1123 (2002).
DR EMBL; AB015532; AAN53942.1; -
DR TIGR; SC00866; -
KW Complete proteome.
SQ SEQUENCE 139 AA; 14811 MW; 41EC1CFA76957920 CRC64;

Query Match 14.2%; Score 110; DB 16; Length 139;
Best Local Similarity 28.3%; Pred. No. 0.13;
Matches 30; Conservative 19; Mismatches 33; Indels 24; Gaps 3;

QY 39 SGPSTSTIYOGSANNALALQSDARKSETTITGSGYGADYDQVTVRVTHMHAFF 98
DB 41 SGRDNLIDVQCGTANCGIYQSSDMS-AVYQAGNDN-----ISLVTOGT----- 87
QY 99 NNATTIDQWNAKNSDITVGQYGGNNAALVNOTASDSSVMVQVQVGN 144

```

Db 88 -----NNBVQLQVGAQKASITQIGNDNLVQLQSGN 122

RESULT 12

Q8EFU3 ID Q8EFU3 PRELIMINARY; PRT; 1422 AA.
AC Q8EFU3;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Lipoprotein, putative.
GN S01872.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.
OX NCBI_TaxID=70863;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NR-1;
RX MEDLINE=2297686; PubMed=12368813;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Mehe B., Clayton R.A.,
RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
RA Vamathevan J., Weidman J., Imprahm M., Lee K., Berry K., Lee C.,
RA Mueller J., Khouri H., Gill J., Uterback T.R., McDonald L.A.,
RA Feldblyum T.V., Smith H.O., Venter J.C., Neilson K.H., Fraser C.M.,
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis.";
RL Nat. Biotechnol. 20:1118-1123 (2002).
DR EMBL; AEO15630; AANS4924.1; -
DR TIGR; S01872; -
DR InterPro; IPR000437; Prok_Lipoprot. S.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KM Complete proteome.
SQ SEQUENCE 1422 AA; 148253 MW; 840DB6300C993DE8 CRC64;

Query Match 13.8%; Score 108; DB 16; Length 1422;
Best Local Similarity 26.5%; Pred. No. 3.4;
Matches 36; Conservative 23; Mismatches 47; Indels 30; Gaps 5;

QY 7 AAFPAIVSGSALAGVVPQWGGGNNHNGSGSPDSTLS---IYQSGANALALQSD 62
DB 1294 AAFPAIFTDGS-----VVTWGGWGDGNGSDSSAVAEQLTSGVKAISTVSAFPAI----- 1343

QY 63 ARKSETTIQSGYGNADYQVLTTRV-----THEMAHFRNATIDQNNAKNSDIT 114
DB 1344 -KDDSSVVTWGGVFWGDSSEVAGQLTSGVKAISTDSAPAAVKDDGLVVTWGSDDN--- 1399

QY 115 VQYGGNNALVNOTA 130
DB 1400 -----GGNSSSVASOLA 1411

RESULT 13
Q92U8 ID Q92U8 PRELIMINARY; PRT; 2174 AA.
AC Q92U8;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical glycine-rich protein Smb21548.
GN RB0989 OR Smb21548.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacterid; PSymb (megaplasmid 2).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;

RX MEDLINE=21396508; PubMed=11481431;
RA Finan T.M., Weidner S., Wong K., Buhrmester J., Chain P.,
RA Vorhoefer F.J., Hernandez-Lucas I., Becker A., Cowie A., Gouzy J.,
RA Golding B., Puehler A.;
RT "The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-
RT fixing endosymbiont Sinorhizobium meliloti.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894 (2001).

DR EMBL; AL603645; CAC49389.1; -
DR PIR; E95965; E95965.
DR CO; GO:0046821; C:extrachromosomal DNA; IEA.
DR GO; GO:0004601; P:peroxidase activity; IEA.
DR GO; GO:0006979; P:response to oxidative stress; IEA.
DR InterPro; IPR005546; Autotransporter.
DR InterPro; IPR002016; Peroxidase.
DR Pfam; PF03797; Autotransporter; 1.
DR PROSITE; PS00435; PEROXIDASE 1; 1.
DR PROSITE; PS00583; PFKB_KINASES_1; 2.
KM Plasmid; Hypothetical protein; Complete proteome.
SQ SEQUENCE 2174 AA; 203314 MW; 008EB68297B44182 CRC64;

Query Match 13.8%; Score 107; DB 16; Length 2174;
Best Local Similarity 26.2%; Pred. No. 6.8;
Matches 37; Conservative 21; Mismatches 61; Indels 22; Gaps 6;

QY 11 AIVSGSALAGVVPQ--WGGGNNHNGSGSPDSTLSIYQSGANALALQSDAR--K 65
DB 693 AIAATGAGAVGILAQSIGGGGNN---GNNATGGDAGFGSFGIAGGGGGGAVANTAVGPK 749

QY 66 SETTIQSGYGNADYQVLTTRVTHEMAHFRNATIDQNNAK--NSDITVQYGVGNN 122
DB 750 GLTITQSGHAAG-----IVAGSGGGGCTGTASSYAGIGCTTASVAVGCTGNG 800

QY 123 AA--LVNOTASDSSVVRQVG 141
DB 801 GAGGVSVLTDSAITGQGG 821

RESULT 14

Q8EIH4 ID Q8EIH4 PRELIMINARY; PRT; 502 AA.
AC Q8EIH4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Conserved hypothetical protein.
GN S00865.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.
OX NCBI_TaxID=70863;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NR-1;
RX MEDLINE=2297686; PubMed=12368813;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Mehe B., Clayton R.A.,
RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
RA Vamathevan J., Weidman J., Imprahm M., Lee K., Berry K., Lee C.,
RA Mueller J., Khouri H., Gill J., Uterback T.R., McDonald L.A.,
RA Feldblyum T.V., Smith H.O., Venter J.C., Neilson K.H., Fraser C.M.,
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis.";
RL Nat. Biotechnol. 20:1118-1123 (2002).
DR EMBL; AEO15532; AANS3941.1; -
DR TIGR; S00865; -
DR InterPro; IPR000437; Prok_Lipoprot. S.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KM Complete proteome.
SQ SEQUENCE 502 AA; 52441 MW; D08CA23D6C46B62D CRC64;

Query Match 13.7%; Score 106.5; DB 16; Length 502;
Best Local Similarity 24.2%; Pred. No. 1.2;

Matches 38; Conservative 25; Mismatches 59; Indels 35; Gaps 4;

```

QY 29 GGNHNGGNSGSPDSTLSTIYOGSANA-----LALSDARKSETTITGSGYNG 78
   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 347 GDNNEIYAFATGEDNSIEISQEGDANFAIVDATGNDNEVVEQDQDQNETITVEG-NNN 405
   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY 79 ADYD-----QLVTRVTHMAHAFRNATIDQ-----NAKNSDIT 114
   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 406 ADVTALQHRGDLNLDLIEEDENAEITQAGSGNWVGSDSSSFAASFGVSGDNNSLM 465
   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY 115 VGOYGNNAALVNOQRSDSSVMVRQVFGNNATANQY 151
   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 466 ITQTGNDNLVIGSQAGNNNSISVTQSGDMNVAIVQY 502
   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

```

RESULT 15

```

ID 089J13      PRELIMINARY;      PRT;      171 AA.
AC 089J13;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE CsgA protein.
GN CsgA OR BIL5300.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RX MEDLINE=22484998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uehimi T.,
RA Sasamoto S., Watanabe A., Ideasa K., Iriuchi M., Kawashima K.,
RA Kohara M., Matsunoto M., Shimo S., Tsuruoka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110.";
RL DNA Res. 9:189-197(2002).
DR EMBL; AP005954; BAC50565.1; -.
KW Complete proteome.
SQ SEQUENCE 171 AA; 17448 MW; 995DB08C01498381 CRC64;

```

Query Match 12.8%; Score 99.5; DB 16; Length 171;

Best Local Similarity 24.7%; Pred. No. 1.2;

Matches 39; Conservative 26; Mismatches 56; Indels 37; Gaps 6;

```

QY 1 MKLKVAFALIVGSLAGVFPWGGGNHNGGNSGPDSTLSTIYOGSANAALALQ 60
   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 40 MKLFFASVAVALSSAAQAA-----NTSTTVQVGLVNGSSVTQ 78
   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY 61 SDARKSETTITGSGYNGADYDQLVTRVTHMAHAFRNATIDQWNAKNSDITVGOYG- 119
   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 79 NGLTNDSSTTQIGLNGASTMGTS-----SPSLNVSTVNOAGVONS-ATTGQYAF 130
   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY 120 GNNALVNOQA-----SDSSVMVRQVFG-NNATANQ 150
   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 131 GNGSAITQNSFCPPALQNNNSASVQQLSFGINTSTVSQ 168
   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

```

Search completed: March 11, 2004, 18:40:47
Job time : 31.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM proteoln - protein search, using sw model

Run on: March 11, 2004, 18:13:53 ; Search time 45.9 Seconds
(without alignments)
929.514 Million cell updates/sec

Title: US-09-543-407-30

Perfect score: 768
Sequence: 1 MLLKVAAPAIIVSGSALA.....DSSVMYRQVFGNNATANDY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp19808:*
2: geneseqp19908:*
3: geneseqp20008:*
4: geneseqp20018:*
5: geneseqp20028:*
6: geneseqp20038:*
7: geneseqp20038:*
8: geneseqp20048:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	768	100.0	151	3	AAB36355
2	760	91.1	151	3	AAB36352
3	681	88.7	151	3	AAB36352
4	681	88.7	151	3	AAB36341
5	676	88.0	151	2	AAW23570
6	659	85.8	151	3	AAB36347
7	657	85.5	151	3	AAB36351
8	653	85.5	151	3	AAB36354
9	603	78.5	151	3	AAB36346
10	600	78.0	151	3	AAB36353
11	599	78.0	151	3	AAB36349
12	597	77.7	151	3	AAB36350
13	566	73.7	151	3	AAB36348
14	517	67.3	151	3	AAB36343
15	512	66.7	151	7	ABR82651
16	496	64.6	120	2	AAW23569
17	496	64.6	120	2	AAW23569
18	439	57.2	142	2	AAW23569
19	367	47.8	122	2	AAW23569
20	215	28.0	45	3	AAB36316
21	132	17.2	22	3	AAB36318
22	115	15.0	22	3	AAB36325
23	115	15.0	22	3	AAB36339
24	115	15.0	22	3	AAB36320
25	113	14.7	24	7	ABR82644

25	109	14.2	23	3	AAB36340	AAB36340
27	109	14.2	23	3	AAB36324	AAB36324
28	109	14.2	23	3	AAB36319	AAB36319
29	104.5	13.6	151	3	AAB36344	AAB36344
30	102	13.3	26	7	ABR82649	ABR82649
31	101.5	13.2	520	6	AAO16497	AAO16497
32	99.5	13.0	151	3	AAB36342	AAB36342
33	98	12.8	26	7	ABR82645	ABR82645
34	96	12.5	19	3	AAB36323	AAB36323
35	96	12.5	19	3	AAB36336	AAB36336
36	96	12.5	19	3	AAB36328	AAB36328
37	95.5	12.4	908	4	ABW65007	ABW65007
38	95	12.4	718	3	AAW78364	AAW78364
39	94.5	12.3	908	4	ABW62355	ABW62355
40	94	12.2	354	7	ABO23520	ABO23520
41	93.5	12.2	2204	6	ABU36440	ABU36440
42	92.5	12.0	738	2	AAW56163	AAW56163
43	92.5	12.0	850	4	ABW65764	ABW65764
44	92.5	12.0	1028	4	ABW62708	ABW62708
45	92.5	12.0	1721	6	ABU34624	ABU34624

ALIGNMENTS

RESULT 1
AAB36355
ID AAB36355 standard; protein; 151 AA.
XX
AC AAB36355;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa::PT3#10 amino acid sequence SEQ ID NO:30.
XX
KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
KW vaccine; immune response; immunogen.
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
PN WO200060102-A2.
XX
XX 12-OCT-2000.
PD
XX
PF 05-APR-2000; 2000MO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
(UYVI-) UNIV VICTORIA.
PI White AP, Doran JL, Collison SK, Kay WM;
WPI; 2000-672631/65.
DR N-PSDB; AAC64631.
XX
PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 139; 139pp; English.
XX
CC The present invention describes a recombinant agfa gene (1) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/7AF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, Caga and Agfa-homologue fimbria subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native

CC copy of that gene; and (4) eliciting an immune response in an animal.
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

SO Sequence 151 AA;

Query Match 100.0%; Score 768; DB 3; Length 151;
 Best Local Similarity 100.0%; Pred. No. 1.7e-66;
 Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLLKVAAPAAIIVSGSALAGVPPQWGGGNNNGSGSPDSTLSIYQGSANALALQ 60
 DB 1 MLLKVAAPAAIIVSGSALAGVPPQWGGGNNNGSGSPDSTLSIYQGSANALALQ 60
 QY 61 SPARKSETTITQSGYNGADVGQADNSTITLTONGFPNNATYDQLVTRVYTHMAHAGG 120
 DB 61 SPARKSETTITQSGYNGADVGQADNSTITLTONGFPNNATYDQLVTRVYTHMAHAGG 120
 QY 121 NNAALVNOTASDSSVWVRQVGFNNATANOY 151
 DB 121 NNAALVNOTASDSSVWVRQVGFNNATANOY 151

RESULT 2

AAB36352 AAB36352 standard; protein; 151 AA.

AC AAB36352;
 XX
 DT 26-FEB-2001 (first entry)
 DE Agfa::PT3#7 amino acid sequence SEQ ID NO:24.
 DE
 KW *Salmonella*: agfa; chromosomal gene replacement; fimbrial; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS *Salmonella enteritidis*.
 OS *Escherichia coli*.
 OS Synthetic.
 OS
 FN WO20060102-A2.
 FN
 PD 12-OCT-2000.
 PD
 PF 05-APR-2000; 2000WO-CA000356.
 PF
 PR 05-APR-1999; 99US-0127888P.
 PR
 PA (UUVI-) UNIV VICTORIA.
 PA
 XX White AP, Doran JL, Collison SK, Kay WW;
 XX WPI; 2000-672631/65.
 DR N-PSDB; AAC64628.
 DR
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 138; 139pp; English.

XX The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF7/RAF) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
 CC Agfa, GsaA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal.
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

SO Sequence 151 AA;

Query Match 91.1%; Score 700; DB 3; Length 151;
 Best Local Similarity 89.9%; Pred. No. 6.8e-60;
 Matches 143; Conservative 0; Mismatches 0; Indels 16; Gaps 2;

QY 1 MLLKVAAPAAIIVSGSALAGVPPQWGGGNNNGSGSPDSTLSIYQGSANALALQ 60
 DB 1 MLLKVAAPAAIIVSGSALAGVPPQWGGGNNNGSGSPDSTLSIYQGSANALALQ 60
 QY 61 SPARKSETTITQSGYNGADVGQADNSTITLTONGFPNNAT-----YDQLVTRVY 112
 DB 61 SPARKSETTITQSGYNGADVGQADNSTITLTONGFPNNATIDQWNAKNYDQLVTRVY 120
 QY 113 HEMAHAGNNAALVNOTASDSSVWVRQVGFNNATANOY 151
 DB 121 HEMAH-----NOTASDSSVWVRQVGFNNATANOY 151

RESULT 3

AAR74625 AAR74625 standard; protein; 151 AA.

AC AAR74625;
 XX
 DT 25-MAR-2003 (revised)
 DT 26-JUN-1995 (first entry)
 DE Agfa sequence.
 DE
 KW *Salmonella*: Agfa; vaccine.
 KW
 OS *Salmonella*.
 OS
 FN WO9425598-A2.
 FN
 PD 10-NOV-1994.
 PD
 PF 26-APR-1994; 94WO-IB000207.
 PF
 PR 26-APR-1993; 93US-00054452.
 PR
 PA (UUVI-) UNIV VICTORIA INNOVATION & DEV CORP.
 PA (KING/) KING J.
 XX

PI Kay WW, Collinson SK, Clouthier SC, Doran JL;
 DR WPI; 1994-358275/44.
 DR N-PSDB; AAC87467.
 XX
 PT Eliciting an immune response to Salmonella - using attenuated Salmonella
 PT strains, vector constructs, or compens. contg. fimbrial type proteins.
 XX
 PS Disclosure; Fig 7B; 95pp; English.
 CC The Salmonella Agfa protein and DNA are used in vaccine and genetic
 CC immunization compositions, respectively, to elicit an immune response to
 CC Salmonella in animals (e.g. food producing animals) and humans. (Updated
 CC on 25-MAR-2003 to correct PN field.)
 CC
 SQ Sequence 151 AA;
 Query Match 88.7%; Score 681; DB 2; Length 151;
 Best Local Similarity 90.7%; Pred. No. 4.7e-58;
 Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;
 QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
 DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
 QY 61 SDARKSETTITGSGYNGADVGGADNSTIELTONGFRNNATYDOLVTRVTHEMAAGG 120
 DB 61 SDARKSETTITGSGYNGADVGGADNSTIELTONGFRNNATIDONNAKNSDITVQYGG 120
 QY 121 NNAALVNQTASDSVWVRQVGFNNATANQY 151
 DB 121 NNAALVNQTASDSVWVRQVGFNNATANQY 151
 RESULT 4
 ID AAB36341 standard; protein; 151 AA.
 XX AAB36341;
 AC
 DT 26-FEB-2001 (first entry)
 DE Salmonella enteritidis Agfa amino acid sequence SEQ ID NO:5.
 XX
 KM Salmonella; agfa; chromosomal gene replacement; fimbrian; epitope;
 KM vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 XX
 PN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000MO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collinson SK, Kay WW;
 DR WPI; 2000-672631/65.
 DR N-PSDB; AAC64617.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 135; 139pp; English.
 CC The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:

CC (1) use of thin aggregative fimbriae (SEF17/7AP) nucleation depended
 CC assembly system of strains of Salmonella, *Escherichia coli* and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrian subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a host cell and introducing the
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrian protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 CC
 SQ Sequence 151 AA;
 Query Match 88.7%; Score 681; DB 3; Length 151;
 Best Local Similarity 90.7%; Pred. No. 4.7e-58;
 Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;
 QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
 DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
 QY 61 SDARKSETTITGSGYNGADVGGADNSTIELTONGFRNNATYDOLVTRVTHEMAAGG 120
 DB 61 SDARKSETTITGSGYNGADVGGADNSTIELTONGFRNNATIDONNAKNSDITVQYGG 120
 QY 121 NNAALVNQTASDSVWVRQVGFNNATANQY 151
 DB 121 NNAALVNQTASDSVWVRQVGFNNATANQY 151
 RESULT 5
 ID AAM23570 standard; protein; 151 AA.
 XX AAM23570;
 AC
 DT 25-MAR-2003 (revised)
 DT 29-SEP-1997 (first entry)
 DE Salmonella enteritidis 27655-3b agfa.
 XX
 KM Enteropathogenic bacteria; enterobacteria; S. enteritidis; antibody.
 XX
 OS Salmonella enteritidis.
 XX
 PN US5635617-A.
 XX
 PD 03-JUN-1997.
 XX
 PF 26-APR-1994; 94US-00233788.
 XX
 PR 26-APR-1993; 93US-00054452.
 XX
 PA (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
 XX
 PI Collinson SK, Kay WW, Doran JL;

XX DR WPI; 1997-309886/28.
 XX DR N-PSDB; AAT74142.
 PT Isolated *Salmonella* gene agfa - used for diagnosis of *Salmonella* or
 PT enteropathogenic bacteria of the Enterobacteriaceae family.
 XX PS Example 2; Fig 7; 85pp; English.
 XX The present sequence represents agfa encoded by the full agfa gene
 CC derived from *Salmonella enteritidis* 27655-3b. The nucleic acid can be
 CC used to provide diagnostic assays for *Salmonella* and/or enteropathogenic
 CC bacteria of the family Enterobacteriaceae. It can also be used to provide
 CC proteins and antibodies which can be used for assays. The nucleic acid
 CC sequence can be used to provide probes or primers which can specifically
 CC hybridize to nucleic acid molecules from greater than 99% of *Salmonella*
 CC strains that are pathogenic to warm-blooded animals relative to nucleic
 CC acid molecules from virtually all other microbial organisms. (Updated on
 XX 25-MAR-2003 to correct PF field.)
 SQ Sequence 151 AA;
 Query Match 88.0%; Score 676; DB 2; Length 151;
 Best Local Similarity 90.1%; Pred. No. 1.4e-57;
 Matches 136; Conservative 2; Mismatches 13; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAIIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANNAALALQ 60
 DB 1 MKLLKVAFAIIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANNAALALQ 60
 QY 61 SPARKSETTITGSGYGNADVGQADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAG 120
 DB 61 SPARKSETTITGSGYGNADVGQADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAG 120
 QY 121 NNAALVNOTASDSSVWVQVGFNNATANOY 151
 DB 121 NNPALVNOTASDSSVWVQVGFNNATANOY 151
 RESULT 6
 AAB36347
 ID AAB36347 standard; protein; 151 AA.
 AC AAB36347;
 XX 26-FEB-2001 (first entry)
 DT 26-FEB-2001 (first entry)
 DE Agfa::PT3#2 amino acid sequence SEQ ID NO:14.
 XX KW *Salmonella*; agfa; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 XX OS *Salmonella enteritidis*.
 OS *Escherichia coli*.
 OS Synthetic.
 XX WO200060102-A2.
 PD 12-OCT-2000.
 XX 05-APR-2000; 2000WO-CA000356.
 PF 05-APR-1999; 99US-0127888P.
 XX (UVT-) UNIT VICTORIA.
 PA White AP, Doran JL, Collison SK, Kay WW;
 PI WPI; 2000-672631/65.
 DR N-PSDB; AAC64623.
 XX PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa

PT protein useful for eliciting immune response in animal.
 XX XX Disclosure; Page 136; 139pp; English.
 XX The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/7AF) nucleation depended
 CC assembly system or strains of *Salmonella*, *Escherichia coli* and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX SQ Sequence 151 AA;
 Query Match 85.8%; Score 659; DB 3; Length 151;
 Best Local Similarity 81.9%; Pred. No. 6.4e-56;
 Matches 136; Conservative 0; Mismatches 0; Indels 30; Gaps 2;
 QY 1 MKLLKVAFAIIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANNAALALQ 60
 DB 1 MKLLKVAFAIIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANNAALALQ 60
 QY 61 SPARKSETTITGSGYGNADVGQADNSTIELTQNGFRNNAT-----YDQ 105
 DB 61 SPARKSETTITGSGYGNADVGQADNSTIELTQNGFRNNATYDQNVAKNSDITVGOYDQ 120
 QY 106 LVTRVVTHEMAHAGGNNALVNOTASDSSVWVQVGFNNATANOY 151
 DB 121 LVTRVVTHEMAHAG-----SVWVQVGFNNATANOY 151
 RESULT 7
 AAB36351
 ID AAB36351 standard; protein; 151 AA.
 AC AAB36351;
 XX 26-FEB-2001 (first entry)
 DT 26-FEB-2001 (first entry)
 DE Agfa::PT3#6 amino acid sequence SEQ ID NO:22.
 XX KW *Salmonella*; agfa; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 XX OS *Salmonella enteritidis*.
 OS *Escherichia coli*.
 OS Synthetic.
 XX WO200060102-A2.
 PD 12-OCT-2000.
 XX 05-APR-2000; 2000WO-CA000356.

PR 05-APR-1999; 99US-0127888P.
 XX (UYVI-) UNIV VICTORIA.
 PA White AP, Doran JL, Collison SK, Kay MW,
 XX WPI; 2000-672631/65.
 DR N-PSDB; AAC64627.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 137, 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 XX Sequence 151 AA:
 SQ
 Query Match 85.5%; Score 657; DB 3; Length 151;
 Best Local Similarity 81.9%; Pred. No. 1e-55;
 Matches 136; Conservative 0; Mismatches 0; Indels 30; Gaps 2;
 QY 1 MKLLKVAAPFAIVVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYOGSANAALALQ 60
 DB 1 MKLLKVAAPFAIVVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYOGSANAALALQ 60
 QY 61 SDARKSETTITGSGYNGADVGAGDNSTIELTONGFRNNATYDQVTRVTHEMAHA-- 118
 DB 61 SDARKSETTITGSGYNGADVGAGDNSTIELTONGFRNNATYDQVTRVTHEMAHAAD 105
 QY 119 -----GGNNAALVNOTASDSSVWVRQVFGNNATANQY 151
 DB 106 WNAKNSDITVGYGGNNALVNOTASDSSVWVRQVFGNNATANQY 151
 RESULT 8
 AAB36354
 ID AAB36354 standard; protein; 151 AA.
 AC AAB36354;
 XX 26-FEB-2001 (first entry)
 XX Agfa: P73#9 amino acid sequence SEQ ID NO:28.
 DE *Salmonella*; agfa; chromosomal gene replacement; fimbrial; epitope;
 XX vaccine; immune response; immunogen.
 KW

XX *Salmonella enteritidis*.
 OS *Escherichia coli*.
 OS Synthetic.
 XX
 PN MO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 XX (UYVI-) UNIV VICTORIA.
 PA White AP, Doran JL, Collison SK, Kay MW,
 PI WPI; 2000-672631/65.
 DR N-PSDB; AAC64630.
 DR
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 138, 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 XX Sequence 151 AA:
 SQ
 Query Match 79.8%; Score 613; DB 3; Length 151;
 Best Local Similarity 74.6%; Pred. No. 1.9e-51;
 Matches 129; Conservative 0; Mismatches 0; Indels 44; Gaps 2;
 QY 1 MKLLKVAAPFAIVVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYOGSANAALALQ 60
 DB 1 MKLLKVAAPFAIVVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYOGSANAALALQ 60
 QY 61 SDARKSETTITGSGYNGADVGAGDNSTIELTONGFRNNATYDQVTRVTHEMAHA-- 118
 DB 61 SDARKSETTITGSGYNGADVGAGDNSTIELTONGFRNNATYDQVTRVTHEMAHAAR 98
 QY 119 -----GGNNAALVNOTASDSSVWVRQVFGNNATANQY 151
 DB 99 NNATIDQWNAKNSDITVGYGGNNALVNOTASDSSVWVRQVFGNNATANQY 151
 RESULT 9

```

AAB36346
ID AAB36346 standard; protein, 151 AA.
XX
AC AAB36346;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa::PT3#1 amino acid sequence SRQ ID NO.12.
XX
KM Salmonella; agfA; chromosomal gene replacement; fimbria; epitope;
KW vaccine; immune response; immunogen.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYV-) UNIV VICTORIA.
PI White AP, Doran JL, Collison SK, Kay MW,
DR WPI; 2000-672631/65.
N-PSDB; AAC64622.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
XX protein useful for eliciting immune response in animal.
XX
Disclousure; Page 135; 139pp; English.
CC
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
Sequence 151 AA:
Query Match 78.5%; Score 603; DB 3; Length 151;
Best Local Similarity 80.8%; Pred. No. 1.7e-50;
Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0.
1 MKLLKTAFAFAAIYSSALAGVVPPMGGGGNHNGGNSGSPDSTLYTGGSANNAALAQ 60
1 MKLLKTAFAFAAIYSSALAGVVPPMGGGGNHNGGNSGSPDSTLYTGGSANNAALAQ 60

```

Oy		61	SDAKSETTTTGGGCGAGVGGAGDNTSTELTONGRNNATADOLVTRVTTHMAHAGS	120
Dd		61	SDARKSETTTTGGGCGAGVGGAGDNTSTELTONGRNNATTIDQWNAKSIDITVGQYGS	120
Oy		121	NNAALVNQTASDSSVMVRQVGFNNATANQY	151
Dd		121	NNAALVVNYDQLVTRVVTTHMAHANNTATANQY	151
RESULT 10				
AAB36353				
ID	AAB36353	standard; protein; 151 AA.		
XX				
AC	AAB36353;			
XX				
DT	26-FEB-2001	(first entry)		
XX				
DE	AgfA::PT3#8	amino acid sequence SEQ ID NO:26.		
KX	Salmonella; agfA; chromosomal gene replacement; fimbria; epitope;			
KW	vaccine; immune response; immunogen.			
XX				
OS	Salmonella enteritidis.			
OS	Escherichia coli.			
XX	Synthetic.			
PN	MO200060102-A2.			
PX				
PD	12-OCT-2000.			
PX				
PF	05-APR-2000; 2000OWO-CA000356.			
PR	05-APR-1999; 99US-0127888P.			
PA	(UYVI-) UNIT VICTORIA.			
PI	White AP, Doran JL, Collison SK, Kay WJ;			
XX				
DR	WPI; 2000-672631/65.			
DR	N-PSDB; AAC64629.			
XX				
PT	Recombinant agfA gene having a segment replaced by a foreign DNA sequence			
PT	which encodes foreign epitope or antigen, expresses recombinant AgfA			
XX	protein useful for eliciting immune response in animal.			
PS	Disclosure; Page 138; 139pp; English.			
CC	The present invention describes a recombinant agfA gene (1) where a			
CC	segment of the gene has been replaced by a segment of a foreign DNA			
CC	sequence which encodes a foreign epitope or antigen. Also described are:			
CC	(1) use of thin aggregative fimbriae (SBFI7/TAF) nucleation depended			
CC	assembly system of strains of Salmonella, Escherichia coli and			
CC	Enterobacteriaceae for the production of fimbriae comprising recombinant			
CC	AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)			
CC	directing recombination of a recombinant gene into the chromosome of the			
CC	homologous species; (3) directing recombination of a recombinant gene			
CC	back into the chromosome of the homologous species, replacing the native			
CC	copy of that gene; and (4) eliciting an immune response in an animal,			
CC	comprising separating an amino acid polymer comprising a recombinant AgfA			
CC	protein containing a replacement segment or segments of foreign amino			
CC	acid sequence or sequences grown on a Salmonella, E. coli or			
CC	Enterobacteriaceae host cell, from the host cell and introducing the			
CC	polymer into the animal in conjunction with a carrier or diluent. (1) is			
CC	useful for the expression of recombinant AgfA protein which is useful for			
CC	eliciting an immune response in an animal. In a fimbrial presentation			
CC	system the heterologous antigens are presented in high numbers (up to			
CC	500,000 copies/cell), the hybrid fimbrial protein possesses both the			
CC	immunogenicity and adhesion properties relevant for an efficient live			
CC	vaccine, the carrier fimbrial subunit proteins are usually strong			
CC	immunogens, which may be important for directing an immune response			
CC	against the inserted epitope, and hybrid fimbriae are easy and			
CC	inexpensive to purify in large amount. The present sequence is given in			

CC the exemplification of the present invention
 XX Sequence 151 AA;
 SQ

Query Match 78.1%; Score 600; DB 3; Length 151;
 Best Local Similarity 81.5%; Pred. No. 3,4e-50;
 Matches 123; Conservative 4; Mismatches 24; Indels 0; Gaps 0;

QY 1 MKLLKVAAPFAIIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYXGSANALALQ 60
 DB 1 MKLLKVAAPFAIIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYXGSANALALYDQ 60

QY 61 SDARKSETTITGSGYNGADVQGGADNSTIELTONGFRNNATYDQLVTRVTHEMAHAG 120
 DB 61 LVTRVTHEMAHAGYNGADVQGGADNSTIELTONGFRNNATIDQWNAKNSDITVQYGG 120

QY 121 NNAALVNOTASDSSVWVROVGFNNATANQY 151
 DB 121 NNAALVNOTASDSSVWVROVGFNNATANQY 151

RESULT 11
 AAB36349
 ID AAB36349 standard; protein; 151 AA.
 AC AAB36349;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE AgfA::PT3#4 amino acid sequence SEQ ID NO:18.
 XX
 KW Salmonella; agfA; chromosomal gene replacement; fimbria; epitope;
 XX vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 PN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000MO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay W;
 XX
 DR WPI; 2000-672631/65.
 XX
 DR N-PSDB; AAC64625.
 XX

Recombinant agfA gene having a segment replaced by a foreign DNA sequence which encodes foreign epitope or antigen, expresses recombinant AgfA protein useful for eliciting immune response in animal.

Disclosure; Page 136; 139pp; English.

The present invention describes a recombinant agfA gene (1) where a segment of the gene has been replaced by a segment of a foreign DNA sequence which encodes a foreign epitope or antigen. Also described are: (1) use of thin aggregative fimbriae (SEF17/7AF) nucleation depended assembly system of strains of Salmonella, Escherichia coli and Enterobacteriaceae for the production of fimbriae comprising recombinant AgfA, CsgA and AgfA-homologue fimbria subunits, respectively; (2) directing recombination of a recombinant gene into the chromosome of the homologous species; (3) directing recombination of a recombinant gene back into the chromosome of the homologous species, replacing the native copy of that gene; and (4) eliciting an immune response in an animal, comprising separating an amino acid polymer comprising a recombinant AgfA protein containing a replacement segment or segments of foreign amino acid sequence or sequences grown on a Salmonella, E. coli or

CC Enterobacteriaceae host cell, from the host cell and introducing the CC polymer into the animal in conjunction with a carrier or diluent. (1) is CC useful for the expression of recombinant AgfA protein which is useful for CC eliciting an immune response in an animal. In a fimbrial presentation CC system the heterologous antigens are presented in high numbers (up to CC 500,000 copies/cell), the hybrid fimbria protein possesses both the CC immunogenicity and adhesion properties relevant for an efficient live CC vaccine, the carrier fimbrial subunit proteins are usually strong CC immunogens, which may be important for directing an immune response CC against the inserted epitope, and hybrid fimbriae are easy and CC inexpensive to purify in large amount. The present sequence is given in CC the exemplification of the present invention

SQ Sequence 151 AA;

Query Match 78.0%; Score 599; DB 3; Length 151;
 Best Local Similarity 81.5%; Pred. No. 4,2e-50;
 Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAAPFAIIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYXGSANALALQ 60
 DB 1 MKLLKVAAPFAIIVSGSALAGVVPQWGGGNNHNGGNSGPDYDQLVTRVTHEMAHALQ 60

QY 61 SDARKSETTITGSGYNGADVQGGADNSTIELTONGFRNNATYDQLVTRVTHEMAHAG 120
 DB 61 SDARKSETTITGSGYNGADVQGGADNSTIELTONGFRNNATIDQWNAKNSDITVQYGG 120

QY 121 NNAALVNOTASDSSVWVROVGFNNATANQY 151
 DB 121 NNAALVNOTASDSSVWVROVGFNNATANQY 151

RESULT 12
 AAB36350
 ID AAB36350 standard; protein; 151 AA.
 AC AAB36350;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE AgfA::PT3#5 amino acid sequence SEQ ID NO:20.
 XX
 KW Salmonella; agfA; chromosomal gene replacement; fimbria; epitope;
 XX vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 PN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000MO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay W;
 XX
 DR WPI; 2000-672631/65.
 XX
 DR N-PSDB; AAC64626.
 XX

Recombinant agfA gene having a segment replaced by a foreign DNA sequence which encodes foreign epitope or antigen, expresses recombinant AgfA protein useful for eliciting immune response in animal.

Disclosure; Page 137; 139pp; English.

The present invention describes a recombinant agfA gene (1) where a segment of the gene has been replaced by a segment of a foreign DNA sequence which encodes a foreign epitope or antigen. Also described are:

CC (1) use of thin aggregative fimbriae (SEF17/TAf) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant AgfA
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

CC Sequence 151 AA:

Query Match 77.7%; Score 597; DB 3; Length 151;
 Best Local Similarity 81.5%; Pred. No. 6.6e-50;
 Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAFAIIVSGSALAGVVPQWGGGNHNGGNSGDPSTLSIYQGSANALALQ 60
 DB 1 MKLLKVAFAIIVSGSALAGVVPQWGGGNHNGGNSGDPSTLSIYQGSANALALQ 60
 QY 61 SDARKSETTITQSGYGNAGADVQGADNSTIELTONGFRNATYDQLVTRVTHMAHAGG 120
 DB 61 SDARKYDQLVTRVTHMAHAGGADNSTIELTONGFRNATYDQMNANKSDITVGYGG 120
 QY 121 NNAALVNOTASDSVVRQVGFNNATANOY 151
 DB 121 NNAALVNOTASDSVVRQVGFNNATANOY 151

RESULT 13
 AAB36348
 ID AAB36348 standard; protein; 151 AA.

AC AAB36348;
 XX 26-FEB-2001 (first entry)
 DE AgfA::PT#3 amino acid sequence SEQ ID NO:16.
 KW *Salmonella*; agfA; chromosomal gene replacement; fimbrial; epitope;
 KM vaccine; immune response; immunogen.
 XX *Salmonella enteritidis*.
 OS *Escherichia coli*.
 OS Synthetic.
 XX WO200060102-A2.
 PN 12-OCT-2000.
 XX 05-APR-2000; 2000WO-CA000356.
 XX 05-APR-1999; 99US-0127888P.
 XX (UYVI-) UNIV VICTORIA.
 XX White AP, Doran JL, Collison SK, Kay WW;
 XX WPI; 2000-672631/65.

DR N-PSDB; AAC64624.
 XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant AgfA
 PT protein useful for eliciting immune response in animal.
 XX Disclosure; Page 136; 139p; English.

CC The present invention describes a recombinant agfA gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAf) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant AgfA
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

CC Sequence 151 AA:

Query Match 73.7%; Score 566; DB 3; Length 151;
 Best Local Similarity 80.8%; Pred. No. 6.7e-47;
 Matches 122; Conservative 5; Mismatches 24; Indels 0; Gaps 0;

QY 1 MKLLKVAFAIIVSGSALAGVVPQWGGGNHNGGNSGDPSTLSIYQGSANALALQ 60
 DB 1 MKLLKVAFAIIVSGSALAGVVPQWGGGNHNGGNSGDPSTLSIYQGSANALALQ 60
 QY 61 SDARKSETTITQSGYGNAGADVQGADNSTIELTONGFRNATYDQLVTRVTHMAHAGG 120
 DB 61 SDARKSETTITQSGYGNAGADVQGADNSTIELTONGFRNATYDQMNANKSDITVGYGG 120
 QY 121 NNAALVNOTASDSVVRQVGFNNATANOY 151
 DB 121 NNAALVNOTASDSVVRQVGFNNATANOY 151

RESULT 14
 AAB36343
 ID AAB36343 standard; protein; 151 AA.

AC AAB36343;
 XX 26-FEB-2001 (first entry)
 DE *Escherichia coli* CsgA amino acid sequence SEQ ID NO:7.
 KW *Salmonella*; agfA; chromosomal gene replacement; fimbrial; epitope;
 KM vaccine; immune response; immunogen.
 XX *Escherichia coli*.
 OS WO200060102-A2.
 PN 12-OCT-2000.

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:25:54 ; Search time 12.4 Seconds
(without alignments)
628,671 Million cell updates/sec

Title: US-09-543-407-30

Sequence: 1 MLLKLVAAALVSGSALA.....DSVWVROYFGNNATANY 151

Scoring table: BLOSUM62

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /cgn2_6/prodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/prodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/prodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/prodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/prodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/prodata/2/iaa/backfileal.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	676	88.0	151	1 US-08-233-788A-59	Sequence 59, App1
2	496	64.6	120	1 US-08-233-788A-57	Sequence 57, App1
3	92.5	12.0	738	3 US-08-864-038A-3	Sequence 5, App1
4	91.5	11.9	673	3 US-09-196-387-8	Sequence 8, App1
5	91.5	11.9	673	4 US-09-841-835-8	Sequence 8, App1
6	91.5	11.9	949	3 US-09-196-387-10	Sequence 10, App1
7	91.5	11.9	949	4 US-09-841-835-10	Sequence 10, App1
8	91.5	11.9	1327	3 US-09-196-387-2	Sequence 2, App1
9	91.5	11.9	1327	4 US-09-841-835-2	Sequence 2, App1
10	91.5	11.9	1327	4 US-09-972-115A-8	Sequence 8, App1
11	86.5	11.3	745	4 US-09-336-115C-6	Sequence 6, App1
12	84	10.9	2736	4 US-09-252-991A-30227	Sequence 30227, A
13	83	10.8	1256	4 US-09-134-000C-5130	Sequence 5130, Ap
14	81.5	10.6	589	4 US-09-489-039A-7849	Sequence 7849, Ap
15	80	10.4	867	4 US-09-540-236-2676	Sequence 2676, Ap
16	78.5	10.2	878	4 US-09-540-236-3401	Sequence 3401, Ap
17	78.5	10.2	309	4 US-09-252-991A-22266	Sequence 22266, A
18	78.5	10.2	873	4 US-09-336-447A-13	Sequence 13, App1
19	78.5	10.2	892	4 US-09-336-447A-5	Sequence 5, App1
20	78.5	10.2	1415	4 US-09-252-991A-26438	Sequence 26438, A
21	77.5	10.1	1160	3 US-08-808-599A-24	Sequence 24, App1
22	77.5	10.1	1207	4 US-09-489-039A-11518	Sequence 11518, A
23	77	10.0	415	4 US-09-025-769B-280	Sequence 280, App
24	77	10.0	975	4 US-09-328-352-4764	Sequence 4764, Ap
25	77	10.0	1003	1 US-08-571-758-4	Sequence 4, App1
26	77	10.0	1003	1 US-08-909-984A-4	Sequence 4, App1
27	77	10.0	1003	1 US-08-909-983-4	Sequence 4, App1

28	77	10.0	1034	4 US-09-252-991A-26658	Sequence 26658, A
29	77	10.0	1690	4 US-09-595-684B-39	Sequence 39, App1
30	77	10.0	1664	2 US-08-804-227C-3	Sequence 3, App1
31	77	10.0	2315	4 US-09-543-681A-5434	Sequence 5434, Ap
32	76.5	10.0	624	4 US-09-336-447A-7	Sequence 7, App1
33	76.5	10.0	941	4 US-09-336-447A-9	Sequence 9, App1
34	76.5	10.0	1398	1 US-08-750-532-9	Sequence 9, App1
35	76.5	10.0	1398	3 US-08-894-818B-8	Sequence 8, App1
36	76.5	10.0	1398	4 US-09-445-472-6	Sequence 6, App1
37	76.5	10.0	1612	1 US-08-169-927-2	Sequence 2, App1
38	76.5	10.0	2123	3 US-08-968-685A-10	Sequence 10, App1
39	76	9.9	231	4 US-09-540-236-3827	Sequence 3827, Ap
40	76	9.9	870	4 US-09-177-650-91	Sequence 91, App1
41	75.5	9.8	238	4 US-09-495-880A-42	Sequence 42, App1
42	75.5	9.8	282	2 US-08-860-174A-10	Sequence 10, App1
43	75.5	9.8	365	4 US-09-489-847-322	Sequence 322, App
44	75.5	9.8	394	4 US-09-252-991A-28148	Sequence 28148, A
45	75.5	9.8	432	4 US-09-403-089A-1	Sequence 1, App1

ALIGNMENTS

RESULT 1
US-08-233-788A-59
Sequence 59, Application US/08233788A

Patent No. 5635617

GENERAL INFORMATION:

APPLICANT: Doran, James L.

APPLICANT: Collinson, Karen S.

APPLICANT: Clouthier, Sharon C.

TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION

NUMBER OF SEQUENCES: 61

CORRESPONDENCE ADDRESS:

ADDRESS: Seed and Berry

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: U.S.A.

ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/233,788A

FILING DATE: 26-APR-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: King, Joshua

REGISTRATION NUMBER: 35,570

REFERENCE/DOCKET NUMBER: 920043.403C2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 682-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 59:

SEQUENCE CHARACTERISTICS:

LENGTH: 151 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-233-788A-59

Query Match

Best Local Similarity 90.1%; Pred. No. 2, 7e-61;

Matches 136; Conservative 2; Mismatches 13; Indels 0; Gaps 0;

QY 1 MLLKLVAAALVSGSALAGVPPWGGGNNNGGNSGPGSTLSIYGSANALALQ 60

Db 1 MLLKVAFAAIVVSGSALAGVVPWGCGGNHNGGNSGPDSTLSIYOXSANALALQ 60
Qy 61 SPARKSETTITGSGYNGADVGGADNSTIELTONGFRNNATYDQLVTRVYTHEMAHAG 120
Db 61 SPARKSETTITGSGYNGADVGGADNSTIELTONGFRNNATIDQMAKNSDITVGYGG 120
Qy 121 NNAALVNOTASDSYVVRQVGFNNATANQY 151
Db 121 NNPALVNOTASDSYVVRQVGFNNATANQY 151

RESULT 2

US-08-233-788A-57
; Sequence 57, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:

APPLICANT: Doran, James L.
APPLICANT: Kay, William W.
APPLICANT: Collinson, Karen S.
APPLICANT: Clouthier, Sharon C.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
TITLE OF INVENTION: OF SALMONELLA
NUMBER OF SEQUENCES: 61
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed and Berry
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: U.S.A.
ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233,788A
FILING DATE: 26-APR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: King, Joshua
REGISTRATION NUMBER: 35,570
REFERENCE/DOCKET NUMBER: 920043.403C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEEDANBERRY
INFORMATION FOR SEO ID NO: 57:
SEQUENCE CHARACTERISTICS:
LENGTH: 120 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-233-788A-57

Query Match 64.6%; Score 496; DB 1; Length 120;
Best Local Similarity 87.5%; Pred. No. 4e-43;
Matches 98; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

Qy 22 VVPWGGGNGNNGGNSGPDSTLSIYOXSANALALQSDARKSETTITGSGYNGADV 81
Db 1 VVPWGGGNGNNGGNSGPDSTLSIYOXSANALALQSDARKSETTITGSGYNGADV 60
Qy 82 GQAGDNSTIELTONGFRNNATYDQLVTRVYTHEMAHAGNNAALVNOTASDS 133
Db 61 GQAGDNSTIELTONGFRNNATIDQMAKNSDITVGYGNNALVNOTASDS 112

RESULT 3

US-08-864-038A-3
; Sequence 3, Application US/08864038A
; Patent No. 6001592
; GENERAL INFORMATION:

APPLICANT: Kunio NAKASHIMA et al.
TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE CDNA, VECTOR
TITLE OF INVENTION: CONTRAINING SAID CDNA, HOST CELLS TRANSFORMED WITH SAID
TITLE OF INVENTION: VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING
TITLE OF INVENTION: SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBODY
TITLE OF INVENTION: TO SAID POLYPEPTIDE
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: 812-5 Hirano
STREET: Ieshinden
CITY: Tsu-city
STATE: Mie-prefecture
COUNTRY: JAPAN
ZIP: 514-01

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: Microsoft Windows 95
SOFTWARE: Word Perfect 6.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/864,038A
FILING DATE: May 28, 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 8-184459
FILING DATE: 15-July-1996
ATTORNEY/AGENT INFORMATION:
NAME: C. Bruce Hamburg
REGISTRATION NUMBER: 22,389
REFERENCE/DOCKET NUMBER: F-5610
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)986-2340
TELEFAX: (212)953-7733
INFORMATION FOR SEO ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 738
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: pinetada fucata
CELL TYPE: mantle epithelial cell
FEATURE:
NAME/KEY: peptide
LOCATION: from 1 to 738
IDENTIFICATION METHOD: E (by experiment)
US-08-864-038A-3

Query Match 12.0%; Score 92.5; DB 3; Length 738;
Best Local Similarity 27.7%; Pred. No. 0.46;
Matches 43; Conservative 12; Mismatches 61; Indels 39; Gaps 6;

Qy 3 LKVAFAAIVVSGSALAGVVPWGCGGNHNGGNSGPDSTLSIYOXSANALALQSD 62
Db 419 LKSHASASASASASASAG-----GGGGGNGGNGGG-----GGAGALA----- 460
Qy 63 ARKSETTITGSGYNGADVGGADNSTIELTONGFRNNATYDQLVTRVYTHEMAHAGNN 122
Db 461 -----AALAAAGAGGGLGGGGGALAAALAAAG-AGGGGCGU-----GSLGGLGGGS 508
Qy 123 AALVNOTASDS-----VVRQVGFNNATA 148
Db 509 AAAAAAASGGRALRRALRRORMGGSAAA 543

RESULT 4

US-09-196-387-8
; Sequence 8, Application US/09196387
; Patent No. 6277613
; GENERAL INFORMATION:
APPLICANT: de Lange, Titia
APPLICANT: Smith, Susan
TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
TITLE OF INVENTION: OF USE THEREOF

NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauder & Jackson
STREET: 411 Hackensack Avenue, 4th Floor
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/196,387
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/095,225
FILING DATE: June 10, 1998
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 673 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-196-387-8

Query Match 11.9%; Score 91.5; DB 3; Length 673;
Best Local Similarity 29.4%; Pred. No. 0.51;
Matches 42; Conservative 17; Mismatches 65; Indels 19; Gaps 6;

QY 6 VAAFAAI-VVSGSALAGVPPQWGGGNNHGGGNSGPDSTLSTIYQGSANAAALALQSDAR 64
DB 99 VAAAPVPAVSTSSAGVAPNPAGSGSNSSPSSPTSS-SSSSPSSGSSLAESPEAA 157
QY 65 KSETTIT---OSGYNGADVCGADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGG 120
DB 158 GVSSTAPLPGAGAGCTGVPVAVSGALRELLACRND-----VSRV--KRLVDAA 206

QY 121 NNAALVNTASDSVWVRQVFG 143
DB 207 VNAK--DMAGRKSSPLHFAAGF 227

RESULT 5
US-09-841-835-8
Sequence 8, Application US/09841835
Patent No. 6506587
GENERAL INFORMATION:
APPLICANT: de Lange, Titia
APPLICANT: Smith, Susan
TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRP1 AND METHODS
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauder & Jackson
STREET: 411 Hackensack Avenue, 4th Floor
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/841,835
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/196,387
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 673 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-841-835-8

Query Match 11.9%; Score 91.5; DB 4; Length 673;
Best Local Similarity 29.4%; Pred. No. 0.51;
Matches 42; Conservative 17; Mismatches 65; Indels 19; Gaps 6;

QY 6 VAAFAAI-VVSGSALAGVPPQWGGGNNHGGGNSGPDSTLSTIYQGSANAAALALQSDAR 64
DB 99 VAAAPVPAVSTSSAGVAPNPAGSGSNSSPSSPTSS-SSSSPSSGSSLAESPEAA 157
QY 65 KSETTIT---OSGYNGADVCGADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGG 120
DB 158 GVSSTAPLPGAGAGCTGVPVAVSGALRELLACRND-----VSRV--KRLVDAA 206

QY 121 NNAALVNTASDSVWVRQVFG 143
DB 207 VNAK--DMAGRKSSPLHFAAGF 227

RESULT 6
US-09-196-387-10
Sequence 10, Application US/09196387
Patent No. 6277613
GENERAL INFORMATION:
APPLICANT: de Lange, Titia
APPLICANT: Smith, Susan
TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRP1 AND METHODS
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauder & Jackson
STREET: 411 Hackensack Avenue, 4th Floor
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/196,387
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/095,225
FILING DATE: June 10, 1998
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.

REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 949 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-196-387-10

Query Match 11.9%; Score 91.5; DB 3; Length 949;
Best Local Similarity 29.4%; Pred. No. 0.8;
Matches 42; Conservative 17; Mismatches 65; Indels 19; Gaps 6;

QY 6 VAAFAAI-VVSGSALAGVVPWGGGNNHNGSGSPDSTLSIYQGSANALALQSDAR 64
DB 99 VAAAPVPAVSTSSAGVAPNPAGSGSNNSPSSSSPTSS-SSSSPSSPGSSLAESPAA 157
QY 65 KSETTIT---OSGYNGADVCGADNSTIELTONGFRNNATYDQLVTRVVTHEMAHAGG 120
DB 158 GVSSTAPLGPAGAGPCTGVPVAVSGALRELLFACRNGD-----VSRV--KRLVDAA 206
QY 121 NNAALVNOTASDSSVWVRQVFG 143
DB 207 VNAK--DMAGRKSSPLHFAAGFG 227

RESULT 7
US-09-841-835-10
Sequence 10, Application US/09841835
Patent No. 6506587
GENERAL INFORMATION:
APPLICANT: de Lange, Titia
APPLICANT: Smith, Susan
TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
TITLE OF INVENTION: OF USE THEREOF
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSER: Klauber & Jackson
STREET: 411 Hackensack Avenue, 4th Floor
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/841,835
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/196,387
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 949 amino acids
TYPE: amino acid
TOPOLOGY: linear

MOLECULE TYPE: protein
US-09-841-835-10

Query Match 11.9%; Score 91.5; DB 4; Length 949;
Best Local Similarity 29.4%; Pred. No. 0.8;
Matches 42; Conservative 17; Mismatches 65; Indels 19; Gaps 6;

QY 6 VAAFAAI-VVSGSALAGVVPWGGGNNHNGSGSPDSTLSIYQGSANALALQSDAR 64
DB 99 VAAAPVPAVSTSSAGVAPNPAGSGSNNSPSSSSPTSS-SSSSPSSPGSSLAESPAA 157
QY 65 KSETTIT---OSGYNGADVCGADNSTIELTONGFRNNATYDQLVTRVVTHEMAHAGG 120
DB 158 GVSSTAPLGPAGAGPCTGVPVAVSGALRELLFACRNGD-----VSRV--KRLVDAA 206
QY 121 NNAALVNOTASDSSVWVRQVFG 143
DB 207 VNAK--DMAGRKSSPLHFAAGFG 227

RESULT 8
US-09-196-387-2
Sequence 2, Application US/09196387
Patent No. 6277613
GENERAL INFORMATION:
APPLICANT: de Lange, Titia
APPLICANT: Smith, Susan
TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
TITLE OF INVENTION: OF USE THEREOF
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSER: Klauber & Jackson
STREET: 411 Hackensack Avenue, 4th Floor
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/196,387
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/095,225
FILING DATE: June 10, 1998
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1327 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-09-196-387-2

Query Match 11.9%; Score 91.5; DB 3; Length 1327;
Best Local Similarity 29.4%; Pred. No. 1.3;
Matches 42; Conservative 17; Mismatches 65; Indels 19; Gaps 6;

QY 6 VAAFAAI-VVSGSALAGVVPWGGGNNHNGSGSPDSTLSIYQGSANALALQSDAR 64

Db 99 VAAAPVPAVTSAGVAPNPAGSGNNSSPSSSPTSS-SSSSPSSPSSLAESPFA 157
Qy 65 KSETTIT-----OSGYNGADVGGAGDNSTIELTONGFRNNATYDQVTRVTHEMAAG 120
Db 158 GVSSTAPLPGAGPGTGVPAVSGALRELLACRNGD-----VSRV--KRLVDAAN 206
Qy 121 NNAALVNQTSADSSVVRQVFG 143
Db 207 VNAK--DMAGRKSSPLHFAAGFG 227

RESULT 9
US-09-841-835-2
; Sequence 2, Application US/09841835
; Patent No. 6506587
; GENERAL INFORMATION:
; APPLICANT: Smith, Susan
; APPLICANT: de Lange, Titia
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; NUMBER OF INVENTION: OF USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/841,835
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/196,387
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1327 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; US-09-841-835-2

Query Match 11.9%; Score 91.5; DB 4; Length 1327;
Best Local Similarity 29.4%; Pred. No. 1.3;
Matches 42; Conservative 17; Mismatches 65; Indels 19; Gaps 6;

Qy 6 VAAFAAI-VVSGALAGVVPQWGGGNNHNGGNSGPDSTLSTIYOGSANAALALQSDAR 64
Db 99 VAAAPVPAVTSAGVAPNPAGSGNNSSPSSSPTSS-SSSSPSSPSSLAESPFA 157
Qy 65 KSETTIT-----OSGYNGADVGGAGDNSTIELTONGFRNNATYDQVTRVTHEMAAG 120
Db 158 GVSSTAPLPGAGPGTGVPAVSGALRELLACRNGD-----VSRV--KRLVDAAN 206
Qy 121 NNAALVNQTSADSSVVRQVFG 143
Db 207 VNAK--DMAGRKSSPLHFAAGFG 227

RESULT 10
US-09-972-115A-8
; Sequence 8, Application US/09972115A
; Patent No. 6599728
; GENERAL INFORMATION:
; APPLICANT: Geron Corporation
; APPLICANT: Gregg, Morin B.
; APPLICANT: Walter, Funk D.
; APPLICANT: Mieczyslaw, Piatyazek A.
; TITLE OF INVENTION: A Second Mammalian Telomerase
; FILE REFERENCE: 080/003C
; CURRENT APPLICATION NUMBER: US/09/972,115A
; PRIOR FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: US 60/128,577
; PRIOR FILING DATE: 2000-04-10
; PRIOR APPLICATION NUMBER: US 60/129,123
; PRIOR FILING DATE: 1999-04-13
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 8
; LENGTH: 1327
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-972-115A-8

Query Match 11.9%; Score 91.5; DB 4; Length 1327;
Best Local Similarity 29.4%; Pred. No. 1.3;
Matches 42; Conservative 17; Mismatches 65; Indels 19; Gaps 6;

Qy 6 VAAFAAI-VVSGALAGVVPQWGGGNNHNGGNSGPDSTLSTIYOGSANAALALQSDAR 64
Db 99 VAAAPVPAVTSAGVAPNPAGSGNNSSPSSSPTSS-SSSSPSSPSSLAESPFA 157
Qy 65 KSETTIT-----OSGYNGADVGGAGDNSTIELTONGFRNNATYDQVTRVTHEMAAG 120
Db 158 GVSSTAPLPGAGPGTGVPAVSGALRELLACRNGD-----VSRV--KRLVDAAN 206
Qy 121 NNAALVNQTSADSSVVRQVFG 143
Db 207 VNAK--DMAGRKSSPLHFAAGFG 227

RESULT 11
US-09-336-115C-6
; Sequence 6, Application US/09336115C
; Patent No. 6576244
; GENERAL INFORMATION:
; APPLICANT: Weitzin, Richard A.
; APPLICANT: Guy, Bruno
; TITLE OF INVENTION: LT and CT in Parenteral Immunization
; FILE REFERENCE: 06132/055002
; CURRENT APPLICATION NUMBER: US/09/336,115C
; PRIOR FILING DATE: 1999-06-18
; PRIOR APPLICATION NUMBER: US 09/100,258
; PRIOR FILING DATE: 1998-06-19
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 745
; TYPE: PRT
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)...(20)
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 721
; OTHER INFORMATION: Xaa = Any Amino Acid
; US-09-336-115C-6

Query Match 11.3%; Score 86.5; DB 4; Length 745;
Best Local Similarity 20.0%; Pred. No. 1.9;
Matches 34; Conservative 27; Mismatches 72; Indels 37; Gaps 5;

QY 8 AFAIIVSGSALAGVPPW-----GGGNNHGGN----- 37
DB 102 AQAQVELAINAAVGL---WNTIGYAVMCGNGTSGPSSVFNPDQGDSTQITCNFE 158
QY 38 SSGPDDTLISYOGSANAALALQSDARKSETTITGSGYGADVGQADNSTIELTONGF 97
DB 159 STPGKSMISDEKKEIKNEAYOIIQALKNQSGPELG-GNGTKV---SYNNYECRQTD 214
QY 98 RNNATYDQLVTRVTHEMAHAGNNALVNQTAASDSSVWVRQVGNNAT 147
DB 215 INGQVQFCXAKKNGSSSSNGNGSSTQTTATTTQDGVITITTYNNKAT 264

RESULT 12

US-09-252-991A-30227
Sequence 30227, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,768
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 30227
LENGTH: 2736
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30227

Query Match 10.9%; Score 84; DB 4; Length 2736;
Best Local Similarity 29.7%; Pred. No. 19;
Matches 43; Conservative 17; Mismatches 49; Indels 36; Gaps 8;

QY 6 VAAFAIIVSGSAL-AGV--PQMGCGNNHGGNSGPDST-----LSIYOGSANA 55
DB 149 MSAGTALLVGAALIGAVAISSSGGG---GGSSVPPDTPPKAASGLKIAPDSSIS 205
QY 56 ALALQSDARKSETTITGSGYGADVGQADNSTIELTONG-----FRNNATYDQLVTR 109
DB 206 GQA-----EAGASVGDITNGDGKPDULVIADANGFTAPLPLNGQTVTV 252
QY 110 VTHEMAHAGNNALVNQTAASDSS 134
DB 253 VVT-----DPAGNASPPAQVTA PDTT 273

RESULT 13

US-09-134-000C-5130
Sequence 5130, Application US/09134000C
Patent No. 6617156
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
FILE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 032796-032
CURRENT APPLICATION NUMBER: US/09/134,000C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/055,778
PRIOR FILING DATE: 1997-08-15
NUMBER OF SEQ ID NOS: 6812
SOFTWARE: PatentIn version 3.1
SEQ ID NO 5130
LENGTH: 1216

TYPE: PRT
ORGANISM: Enterococcus faecalis
US-09-134-000C-5130

Query Match 10.8%; Score 83; DB 4; Length 1216;
Best Local Similarity 24.4%; Pred. No. 8.2;
Matches 49; Conservative 19; Mismatches 55; Indels 78; Gaps 10;

QY 6 VAAFAIIV--SSSALAGVPPWGG--GGNNHGGNSGPDST-----LSIY- 48
DB 181 VTGNTSLVWVDSGRIASI---YGGGYGTATVATVATGVSATKVAITNAATGFOISTY 237
QY 49 ---OYGSANALALQSDARKSETTITGSGYGADVGQ-----GADNSTIELQ 94
DB 238 GGVQYGNIG-----GKVTINDISGGRYTTAGQFFIGSSSRGDIITNATDGIIT 286
QY 95 N-----GFRNNATYDQLVTRVTHEMAHAG-----GNNALVNQTA 130
DB 287 NLNTQLYSAGRADPEGNGQSGTIIIGDITNVVYTAGTNSAGINDFNGAGNNVSKNKS- 345
QY 131 SDSSVWVRQVFGNNATANOY 151
DB 346 -----QIGASNEATYDAY 358

RESULT 14

US-09-489-039A-7849
Sequence 7849, Application US/09489039A
Patent No. 6610836
GENERAL INFORMATION:
APPLICANT: Gary Breton et. al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
FILE REFERENCE: 2709.2004001
CURRENT APPLICATION NUMBER: US/09/489,039A
CURRENT FILING DATE: 2000-01-27
PRIOR APPLICATION NUMBER: US 60/117,747
PRIOR FILING DATE: 1999-01-29
NUMBER OF SEQ ID NOS: 14342
SEQ ID NO 7849
LENGTH: 589
TYPE: PRT
ORGANISM: Klebsiella pneumoniae
US-09-489-039A-7849

Query Match 10.6%; Score 81.5; DB 4; Length 589;
Best Local Similarity 24.5%; Pred. No. 4.5;
Matches 35; Conservative 21; Mismatches 56; Indels 31; Gaps 7;

QY 29 GGNHGGNSGPDSTLSIYOGS-----ANALALQSDARKSETTITGSGYGNGA- 79
DB 259 GGLDRNGANANGQDTDFGIYAFDULTLTERIEINGLALDHYHTKYDSATACGSGRGAI 318
QY 80 --DVQ--GADNSTIELQNG--FRNNATYDQLVTRVTHEMAHAGN--NALVNO 128
DB 319 ACPPOGSGSPVYTYDTRKSGNLVWVKAGALYR-----LITEQ-----GNYVVAISQ 367
QY 129 TASDSSVWVRQVFGNNATANOY 151
DB 368 PGGSSFPALAAAGSGNSANRITDF 390

RESULT 15

US-09-540-236-2676
Sequence 2676, Application US/09540236
Patent No. 6673910
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATA
FILE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 2709.2005-001
CURRENT APPLICATION NUMBER: US/09/540,236
CURRENT FILING DATE: 2000-04-04

THIS PAGE BLANK (USPTO)

Db 78 GNSNSV-----GRDIQKQSGAGNSAIFQEGTGSDELQQTGTSGAVPSGNNMTN 129
Qy 104 DQLVTRVYTHEMAHAGNNALVNGTRASDSVAVRQVGGNNATNQ 150
Db 130 DPGVFNKITOD-SSNSGKSVI-0QKNNVFSIKQNTGNTSTVNO 174

RESULT 2
US-09-820-843A-21
; Sequence 21, Application US/09820843A
; Publication No. US20030039963A1
; GENERAL INFORMATION:
; APPLICANT: Council of Scientific and Industrial Research
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEI
; FILE REFERENCE: 063915
; CURRENT APPLICATION NUMBER: US/09/820,843A
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 21
; LENGTH: 354
; TYPE: PRT
; ORGANISM: M. tuberculosis
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: PPE
; NAME/KEY: misc.feature
; OTHER INFORMATION: gi|1781260
US-09-820-843A-21

Query Match 12.2%; Score 94; DB 10; Length 354;
Best Local Similarity 25.0%; Pred. No. 0.28;
Matches 35; Conservative 16; Mismatches 71; Indels 18; Gaps 5;

Qy 26 MCGGGHNGGNSGSPDS---TLSTYOGSANAALALQSDARKSETT-----TOSGY 75
Db 148 MNGSGTNTGLANAGAGNTGFPDAGNYNPGSLNAGNINSFNSGSGNGLNAGDVNSGV 207
Qy 76 GNGADVQOGADNSTIELTQNGFRNNATYDQLVTRVYTHEMAHAG-----GNNALVNGT 129
Db 208 GNAAGVNTGLGNSG-NINTGGP-NGTINTGTFPSAMTQAGFNSGFNAAGTNSGFGHNDP 265
Qy 130 ASDSSVMTRQVGFNNATN 149
Db 266 AGSGNSGIQNSGFGNSGYV 285

RESULT 3
US-09-841-835-8
; Sequence 8, Application US/09841835
; Patent No. US20020076795A1
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/841,835
; FILING DATE:

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/196,387
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 673 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-841-835-8

Query Match 11.9%; Score 91.5; DB 9; Length 673;
Best Local Similarity 29.4%; Pred. No. 1.1;
Matches 42; Conservative 17; Mismatches 65; Indels 19; Gaps 6;

Qy 6 VAAFAAI-VVSGALAGVYPOWGGGNNHNGGNSGSPSTLSTYOGSANAALALQSDAR 64
Db 99 VAAAPVPAVSTSSAAGVAPNPAGSGNSNPSSSPSS-SSSPSSPGSLASPPAA 157
Qy 65 KSETTIT---OSGYNGADVQOGADNSTIELTQNGFRNNATYDQLVTRVYTHEMAHAG 120
Db 158 GVSSTAPLPGGAAGGTGVPAVSGALRELLACRGD-----VSRV--KRLVDAN 206
Qy 121 NNAALVNGRASSVMNQVGRG 143
Db 207 VNAK--DNAGRRSPLHFAAGFG 227

RESULT 4
US-09-841-835-10
; Sequence 10, Application US/09841835
; Patent No. US20020076795A1
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/841,835
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/196,387
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684

;; PRIOR FILING DATE: 2000-06-28
;; PRIOR APPLICATION NUMBER: 60/141,582
;; PRIOR FILING DATE: 1999-06-29
;; NUMBER OF SEQ ID NOS: 178
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 4
;; LENGTH: 1327
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-10-199-937-4

Query Match 11.3%; Score 91.5; DB 14; Length 1327;
Best Local Similarity 29.4%; Pred. No. 2.6;
Matches 42; Conservative 17; Mismatches 65; Indels 19; Gaps 6;

QY 6 VAAFAAI-VYSGALAGVVPWGGGNNHGGNGSGPDSLTITVYGSANALALQSDAR 64
DB 99 VAAAPVPAVSTSAAGVAPNPAGSGSNNSPSSSPTSS-SSSGSPSSPSSSLAESPEAA 157
QY 65 KSETTIT---OSGYNGADVCGADNSTIELTONGFRNNATYDQLVTRVYTHEMAHAG 120
DB 158 GVSSTAPLPGGAAGPCTGVPAVSGALRELEACRNGD-----VSRV--KRLVDAN 206
QY 121 NNAAVNOTASDSVWVROVFG 143
DB 207 VNAK--DMAGRKSSPLHFAAGFG 227

RESULT 8

US-10-311-406-2
;; Sequence 2, Application US/10311406
;; Publication No. US20030134388A1
;; GENERAL INFORMATION:
;; APPLICANT: Sasaki, Masahiro
;; APPLICANT: Yamada, Hideyuki
;; APPLICANT: Osada, Katsue
;; APPLICANT: Terada, Satoshi
;; TITLE OF INVENTION: MEDIUM SUPPLEMENT AND ANIMAL CELL CULTURE MEDIUM
;; FILE REFERENCE: U 014351-5
;; CURRENT APPLICATION NUMBER: US/10/311,406
;; CURRENT FILING DATE: 2002-12-17
;; PRIOR APPLICATION NUMBER: PCT/JPO2/03836
;; PRIOR FILING DATE: 2002-04-17
;; NUMBER OF SEQ ID NOS: 13
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 2
;; LENGTH: 1217
;; TYPE: PRT
;; ORGANISM: Bombyx Mori
US-10-311-406-2

Query Match 11.5%; Score 88.5; DB 14; Length 1217;
Best Local Similarity 22.2%; Pred. No. 4.6;
Matches 38; Conservative 29; Mismatches 63; Indels 41; Gaps 7;

QY 15 SGSAIAGVPPWGGGNNHGGNGSGPDSLTITVYGSANALAL 58
DB 715 AGSSTSGGTSTYGSRRHGSVSTSSSNTDSSTKNAGSRISGCTYGYSSHRGGS 774
QY 59 LOSDARKSET-----TIT---OSGYNGADVCGADNSTIELLTON-GRPN 100
DB 775 VSSSTGSSSNTDSSTKNAGSRISGCTYGYSSHRGGSVSTSSSNTDSSTKNAGSRIS 834
QY 101 ATYQOLVTRVYTHEMAHAGNNALVNOTASDSVWVROVFGNNATYANCY 151
DB 835 GG-----TSTYGYSSHRGGSVSTSSSNTDS--TKNAGSSTSGSSSTY 878

RESULT 9

US-09-793-306-146
;; Sequence 146, Application US/09793306
;; Patent No. US20020098200A1
;; GENERAL INFORMATION:

;; APPLICANT: Campos-Neto, Antonio
;; APPLICANT: Skeiky, Yasir
;; APPLICANT: Owendale, Pamela
;; APPLICANT: Jen, Shyian
;; APPLICANT: Lodes, Michael
;; APPLICANT: Corixa Corporation
;; TITLE OF INVENTION: Compounds and Methods for Diagnosis and Immunotherapy
;; FILE REFERENCE: 014058-008740US
;; CURRENT APPLICATION NUMBER: US/09/793,306
;; CURRENT FILING DATE: 2001-02-26
;; PRIOR APPLICATION NUMBER: US 60/185,037
;; PRIOR FILING DATE: 2000-02-25
;; PRIOR APPLICATION NUMBER: US 60/223,828
;; PRIOR FILING DATE: 2000-08-08
;; NUMBER OF SEQ ID NOS: 164
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 146
;; LENGTH: 597
;; TYPE: PRT
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: Description of Artificial Sequence:MTTC#3-His
US-09-793-306-146

Query Match 11.3%; Score 87; DB 9; Length 597;
Best Local Similarity 24.2%; Pred. No. 2.7;
Matches 30; Conservative 18; Mismatches 50; Indels 26; Gaps 4;

QY 26 WGGGNNHGGNGSGPDSLTITVYGSANALALQSDARKSETTTTGGYNGADVCGA 85
DB 358 FGSGNNNTGFPNNG-NNNVGFNNGNN-----FPGNAGDINTGF 398
QY 86 DNSTIELTONGFRNNATYDQLVTRVYTHEMAHAGNNALVNOTASDSVWVROVFGNN 145
DB 399 GNAGD--TWTGFGNAGFNMGIGNAGNEDMGVNGSGFNVGAVNAGNS-----VGFENA 451
QY 146 ATAN 149
DB 452 GTLN 455

RESULT 10

US-08-834-666A-6
;; Sequence 6, Application US/08834666A
;; Publication No. US20020044949A1
;; GENERAL INFORMATION:
;; APPLICANT: Kleantous, Harold
;; APPLICANT: Lissolo, Ling
;; APPLICANT: Tomb, Jean-Francois
;; APPLICANT: Miller, Charles
;; APPLICANT: Al-Garawi, Amal
;; TITLE OF INVENTION: 76 Kda Helicobacter Polypeptides and
;; TITLE OF INVENTION: Corresponding Polynucleotide Molecules
;; NUMBER OF SEQUENCES: 64
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Clark & Ribing LLP
;; STREET: 176 Federal Street
;; CITY: Boston
;; STATE: MA
;; COUNTRY: USA
;; ZIP: 02110
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Diskette
;; COMPUTER: IBM Compatible
;; OPERATING SYSTEM: DOS
;; SOFTWARE: FastSeq for Windows Version 2.0
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/834,666A
;; FILING DATE: 01-APR-1997
;; CLASSIFICATION: 424
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Clark, Paul T.

REGISTRATION NUMBER: 30,175
 REFERENCE/DOCKET NUMBER: 06132/038001
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 617-428-0200
 TELEFAX: 617-428-7045
 TELEX:
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 745 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 FRAGMENT TYPE: internal
 FEATURE:
 NAME/KEY: Signal Sequence
 LOCATION: 1...20
 OTHER INFORMATION:
 US-08-834-666A-6

Query Match 11.3%; Score 86.5; DB 8; Length 745;
 Best Local Similarity 20.0%; Pred. No. 4;
 Matches 34; Conservative 27; Mismatches 72; Indels 37; Gaps 5;

QY 8 AFAIVSGSALAGVPPQW-----GGGNNHNGGN----- 37
 DB 102 AYGAVALAINAAVGL--NWTIGYAVWCGNGNGTSGSGSVIFPNDDPGDPTOITCNRF 158
 QY 38 SSGPDSLTISYOGSANAALALQSDARKSETTITSGGCGADVGCGADNSTIELTQNGF 97
 DB 159 STPGKSMSTIDFKLNKAQITIQALKNOSGFPELQ-GNGTIV--SVVYNYECROTAD 214
 QY 98 RNNATYDQLVTRVVTTHMAHAGNNAALVQOTASDSSVWVRQVFGNNAT 147
 DB 215 INGVVQFCARKNKSSSSNGSGSTQTATTGCGVITTTTNNKAT 264

RESULT 11
 US-10-374-780A-590
 Sequence 590, Application US/10374780A
 Publication No. US20040019927A1
 GENERAL INFORMATION:
 APPLICANT: Sherman, Bradley K
 APPLICANT: Riechmann, Jose Luis
 APPLICANT: Jiang, Cai-Zhong
 APPLICANT: Heard, Jacqueline E
 APPLICANT: Haake, Volker
 APPLICANT: Creelman, Robert A
 APPLICANT: Ratcliffe, Oliver
 APPLICANT: Adam, Luc J
 APPLICANT: Reuber, T. Lynne
 APPLICANT: Keddie, James
 APPLICANT: Brown, Pierre E
 APPLICANT: Pilgrim, Marsha L
 APPLICANT: Dubell III, Arnold T
 APPLICANT: Pineda, Omalta
 APPLICANT: Yu, Guo-Liang
 TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES IN PLANTS
 FILE REFERENCE: MBI-0047 CIP
 CURRENT APPLICATION NUMBER: US/10/374,780A
 CURRENT FILING DATE: 2003-02-25
 PRIOR APPLICATION NUMBER: 09/837,944
 PRIOR FILING DATE: 2001-04-18
 PRIOR APPLICATION NUMBER: 60/310,847
 PRIOR FILING DATE: 2001-08-09
 PRIOR APPLICATION NUMBER: 09/934,455
 PRIOR FILING DATE: 2001-08-22
 PRIOR APPLICATION NUMBER: 60/336,049
 PRIOR FILING DATE: 2001-11-19
 PRIOR APPLICATION NUMBER: 60/338,692
 PRIOR FILING DATE: 2001-12-11
 PRIOR APPLICATION NUMBER: 10/171,468
 PRIOR FILING DATE: 2002-06-14

PRIOR APPLICATION NUMBER: 10/225,066
 PRIOR FILING DATE: 2002-08-09
 PRIOR APPLICATION NUMBER: 10/225,067
 PRIOR FILING DATE: 2002-08-09
 PRIOR APPLICATION NUMBER: 10/225,068
 PRIOR FILING DATE: 2002-08-09
 NUMBER OF SEQ ID NOS: 2906
 SOFTWARE: Patentin version 3.2
 SEQ ID NO 590
 LENGTH: 321
 TYPE: PRT
 ORGANISM: Oryza sativa
 FEATURE:
 OTHER INFORMATION: G3384 Orthologous to G256
 US-10-374-780A-590

Query Match 11.1%; Score 85; DB 15; Length 321;
 Best Local Similarity 26.3%; Pred. No. 2;
 Matches 30; Conservative 15; Mismatches 41; Indels 28; Gaps 4;

QY 10 AAIWVSGSALAGVPPQW---GGGNNHNGGNSGPDSTLSIYOGSANAALALQSDARKS 66
 DB 199 AAYASADNIARLLQGMKPGG---GGNGKGPAS-----GSTTTATTQOQPGCS 248
 QY 67 ETTITSGYGGADVGCGADNSTIELTQNGFRNNATYDQLVTRVVTHERAHAGG 120
 DB 249 -----GEGAASASASASOGAAATAATQTECSTETSKMATGG 287

RESULT 12
 US-10-021-811-26
 Sequence 26, Application US/10021811
 Publication No. US20030024007A1
 GENERAL INFORMATION:
 APPLICANT: Cahoon, Rebecca E.
 APPLICANT: Fang, Yiwen
 APPLICANT: Odell, Joan
 APPLICANT: Weng, Zude
 TITLE OF INVENTION: Plant Myb Transcription Factor Homologs
 FILE REFERENCE: BBI294 US NA
 CURRENT APPLICATION NUMBER: US/10/021,811
 CURRENT FILING DATE: 2001-12-14
 PRIOR APPLICATION NUMBER: 60/110,609
 PRIOR FILING DATE: 1998-December-02
 NUMBER OF SEQ ID NOS: 63
 SOFTWARE: Microsoft Office 97
 SEQ ID NO 26
 LENGTH: 323
 TYPE: PRT
 ORGANISM: Oryza sativa
 US-10-021-811-26

Query Match 11.1%; Score 85; DB 14; Length 323;
 Best Local Similarity 26.3%; Pred. No. 2;
 Matches 30; Conservative 15; Mismatches 41; Indels 28; Gaps 4;

QY 10 AAIWVSGSALAGVPPQW---GGGNNHNGGNSGPDSTLSIYOGSANAALALQSDARKS 66
 DB 199 AAYASADNIARLLQGMKPGG---GGNGKGPAS-----GSTTTATTQOQPGCS 248
 QY 67 ETTITSGYGGADVGCGADNSTIELTQNGFRNNATYDQLVTRVVTHERAHAGG 120
 DB 249 -----GEGAASASASASOGAAATAATQTECSTETSKMATGG 287

RESULT 13
 US-10-238-075-749
 Sequence 749, Application US/10238075
 Publication No. US20030148324A1
 GENERAL INFORMATION:
 APPLICANT: I. N. S. E. R. M.
 TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isolated
 TITLE OF INVENTION: E. coli, and biological uses of these polynucleotides and of their

FILE REFERENCE: BLANDINE
 CURRENT APPLICATION NUMBER: US/10/238,075
 CURRENT FILING DATE: 2002-09-10
 PRIOR APPLICATION NUMBER: 0003145
 PRIOR FILING DATE: 2000-03-10
 NUMBER OF SEQ ID NOS: 1576
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO 749
 LENGTH: 1778
 TYPE: PRT
 ORGANISM: Escherichia coli
 US-10-238-075-749

Query Match 11.1%; Score 85; DB 14; Length 1778;
 Best Local Similarity 25.4%; Pred. No. 17;
 Matches 46; Conservative 22; Mismatches 73; Indels 40; Gaps 8;

QY 5 KVAAPAIYVSGSALAGVVPW---GGGNNNGGNSGDPSTLSTIYQ---GSANAL 57
 DB 906 KISSNSTADINSGOLYVADSESYLGGADISDTGVLSGPTVITIGTDYTVGDLAAI 965
 QY 58 -----ALQSPARKSETTITQSYGNG---ADVQAGADNSTIELTONGFRNNATY 103
 DB 966 NTSFSTSLGDALIMDAITAKES-AKHGINNAPSVTITDVANGAVSSTSDAINSGOLYGV 1024
 QY 104 DQLVLT-----RVVTHEMAHAGG--NNAA---LVNOTASDSSVWVRQVGFNN 145
 DB 1025 DYIAALAGNAVNTDGSITPTVYAIAGSYNNVGDALEIDTLDLMDTTANGNG 1084
 QY 146 A 146
 DB 1085 A 1085

RESULT 14
 US-09-905-176-24
 Sequence 24, Application US/09905176
 Patent No. US20020150906A1
 GENERAL INFORMATION:
 APPLICANT: CALIFORNIA INSTITUTE OF TECHNOLOGY
 APPLICANT: Debe, Derek A.
 TITLE OF INVENTION: METHOD FOR DETERMINING THREE-DIMENSIONAL PROTEIN STRUCTURE FROM F
 FILE REFERENCE: 265/297
 CURRENT APPLICATION NUMBER: US/09/905,176
 CURRENT FILING DATE: 2002-04-05
 PRIOR APPLICATION NUMBER: US 60/218,016
 PRIOR FILING DATE: 2000-07-12
 NUMBER OF SEQ ID NOS: 26
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO 24
 LENGTH: 322
 TYPE: PRT
 ORGANISM: Klebsiella pneumoniae
 US-09-905-176-24

Query Match 11.0%; Score 84.5; DB 9; Length 322;
 Best Local Similarity 23.0%; Pred. No. 2.2;
 Matches 43; Conservative 20; Mismatches 55; Indels 69; Gaps 8;

QY 12 IVVGSALAGVPMQGGGNNNGGNSGDPSTLSTIYQGSANAAALQSPARKSETTIT 71
 DB 88 VVIYVTSWTVLPERGGD-----TYSQDN---FLQSRANGVATYRN 125
 QY 72 QSGYG-----NGADVQAGADNS---TIELTONGFRNNATYD----- 104
 DB 126 SDFGLVGLNPALOYGNGSVSGEATNNRGALTKONGDGFSTVYTDLPDGISAGFAY 185
 QY 105 -----QLVTRVVTHEMAHAGG-----NNAAVNO--TASDSSVWVRQVGFNN-- 146
 DB 186 ANSKRTDDONQLLGEGDHAETTYGGLKADANNIYATQYQTVNATRAAGSLGFAKXQON 245
 QY 147 --TANQY 151

DB 246 FEVAQY 252

RESULT 15
 US-09-996-194-16
 Sequence 16, Application US/09996194
 Patent No. US20020151696A1
 GENERAL INFORMATION:
 APPLICANT: Bandaru, Rajasehkar
 TITLE OF INVENTION: 84242, 8035, 55304, 52999, and 21999,
 FILE REFERENCE: No. US20020151696A1 Human Proteins and Methods of Use Thereof
 CURRENT APPLICATION NUMBER: US/09/996,194
 CURRENT FILING DATE: 2001-11-28
 PRIOR APPLICATION NUMBER: 60/250,348
 PRIOR FILING DATE: 2000-11-30
 PRIOR APPLICATION NUMBER: 60/250,073
 PRIOR FILING DATE: 2000-11-30
 PRIOR APPLICATION NUMBER: 60/253,878
 PRIOR FILING DATE: 2000-11-29
 PRIOR APPLICATION NUMBER: 60/250,338
 PRIOR FILING DATE: 2000-11-30
 NUMBER OF SEQ ID NOS: 22
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 16
 LENGTH: 65
 TYPE: PRT
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: Pfam consensus sequence
 US-09-996-194-16

Query Match 10.9%; Score 84; DB 9; Length 65;
 Best Local Similarity 67.9%; Pred. No. 0.33;
 Matches 19; Conservative 3; Mismatches 4; Indels 2; Gaps 1;

QY 98 RNNATYDQLVTRVVTHEMAHAGGNNAAAL 125
 DB 34 RNN--YDQLVTRVVTHEMAHAGGFSVGL 59

Search completed: March 11, 2004, 19:18:41
 Job time : 25.6 secs

Db 121 NNAALVNOTASDSSVWVRQVGFNNATANOY 151

RESULT 2

Major curlin chain precursor (imported) - Salmonella enterica subsp. enterica serovar Typhimurium

A:Species: Salmonella enterica subsp. enterica serovar Typhimurium

A:Note: this species has also been called Salmonella typhimurium

C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002

C:Accession: A10635

R:Parthill, J.; Dongan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Comerford, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Mole, S.; O'Garra, P.

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A:Reference genome sequence of a multiple drug resistant Salmonella enterica serovar Typhimurium

A:Reference number: AB0502; MUID:21534947; PMID:11677608

A:Accession: A10635

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-151 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD08268.1; PID:g16502315; GSPDB:GN00176

C:Genetics:

A:Gene: STY181

Query Match

Best Local Similarity 88.7%; Score 681; DB 2; Length 151;

Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

Db 1 MMLKVAFAIIVVSGSALAGVVPQWGGGNNHGGSGSPDSTLTIYOGSANAALALQ 60

Db 1 MMLKVAFAIIVVSGSALAGVVPQWGGGNNHGGSGSPDSTLTIYOGSANAALALQ 60

Qy 61 SPARKSETTITGSGYGNGADVQGGADNSTIELTONGFNNATYDQVTRVTHMAHAG 120

Db 61 SPARKSETTITGSGYGNGADVQGGADNSTIELTONGFNNATYDQVTRVTHMAHAG 120

Qy 121 NNAALVNOTASDSSVWVRQVGFNNATANOY 151

Db 121 NNAALVNOTASDSSVWVRQVGFNNATANOY 151

RESULT 3

curlin protein csgA precursor - Escherichia coli (strain K-12)

N:Alternate names: csgA protein; major curlin protein

C:Species: Escherichia coli

C:Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 01-Mar-2002

C:Accession: S70788; G64846; S31202; S34560; S34559

R:Hammar, M.; Arngvist, A.; Bian, Z.; Olsen, A.; Normark, S.

Mol. Microbiol. 18, 661-670, 1995

A:Title: Expression of two csg operons is required for production of fibronectin- and CgA:Reference number: S70783; MUID:96414468; PMID:8817489

A:Accession: S70788

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-151 <HAM>

A:Cross-references: EMBL:X90754; NID:g1147558; PIDN:CAA62282.1; PID:g1147564

A:Experimental source: strain K12, substrain W3110

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: G64846

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-151 <BLAT>

A:Cross-references: GB:AB000205; GB:U00096; NID:g1787265; PIDN:AACT4126.1; PID:g1787279;

A:Experimental source: strain K-12, substrain MG1655

R:Olsen, A.; Arngvist, A.; Hammar, M.; Sukupolvi, S.; Normark, S.

Mol. Microbiol. 7, 523-536, 1993

A:Title: The Rpos sigma factor relieves H-NS-mediated transcriptional repression of csg

A:Reference number: S31202; MUID:93211294; PMID:8459772

A:Accession: S31202

A:Molecule type: DNA

A:Residues: 1-6, 'V', 8-151 <OLS>

A:Cross-references: EMBL:L04979

A:Accession: S34560

A:Molecule type: protein

A:Residues: 21-42/44-50 <OLS>

R:Olsen, A.N.; Arngvist, A.M.

A:Note: curli are thin, coiled fibers expressed on the surface of Escherichia coli that

A:Reference number: S34559

A:Accession: S34559

A:Molecule type: DNA

A:Residues: 1-133, 'KQSDSGWV' <OLS>

A:Cross-references: EMBL:L04979; NID:g290424; PIDN:AAA23616.1; PID:g290425

A:Experimental source: strain K-12, substrain W3110

C:Genetics:

A:Gene: csgA

A:Map position: 23.15

C:Function:

A:Description: major component of wild-type curli; interaction between CsgA and CsgB tr

A:Note: curli are thin, coiled fibers expressed on the surface of Escherichia coli that

F:1-20/Domain: signal sequence #status predicted <SIG>

F:21-151/Product: curlin #status experimental <MAT>

Query Match

Best Local Similarity 67.3%; Score 517; DB 2; Length 151;

Matches 105; Conservative 17; Mismatches 29; Indels 0; Gaps 0;

Db 1 MMLKVAFAIIVVSGSALAGVVPQWGGGNNHGGSGSPDSTLTIYOGSANAALALQ 60

Db 1 MMLKVAFAIIVVSGSALAGVVPQWGGGNNHGGSGSPDSTLTIYOGSANAALALQ 60

Qy 61 SPARKSETTITGSGYGNGADVQGGADNSTIELTONGFNNATYDQVTRVTHMAHAG 120

Db 61 SPARKSETTITGSGYGNGADVQGGADNSTIELTONGFNNATYDQVTRVTHMAHAG 120

Qy 121 NNAALVNOTASDSSVWVRQVGFNNATANOY 151

Db 121 NNAALVNOTASDSSVWVRQVGFNNATANOY 151

RESULT 4

curlin major subunit CsgA (imported) - Escherichia coli (strain O157:H7, substrain R1MD

C:Species: Escherichia coli

C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001

C:Accession: D90806

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G

Genesara, N.; Yasunaga, T.; Kihara, S.; Shiba, T.; Hatori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen

A:Reference number: A96293; MUID:21156231; PMID:11258796

A:Accession: D90806

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-152 <HAY>

A:Cross-references: GB:BA000007; PIDN:BA034843.1; PID:g13360880; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain R1MD 0509952

C:Genetics:

A:Gene: Ecs1420

Query Match

Best Local Similarity 65.4%; Score 502.5; DB 2; Length 152;

Matches 104; Conservative 18; Mismatches 29; Indels 1; Gaps 1;

Qy 1 MMLKVAFAIIVVSGSALAGVVPQWGGGNNHGGSGSPDSTLTIYOGSANAALALQ 59

Db 1 MMLKVAFAIIVVSGSALAGVVPQWGGGNNHGGSGSPDSTLTIYOGSANAALALQ 60

Qy 60 QSDARKSETTITGSGYGNGADVQGGADNSTIELTONGFNNATYDQVTRVTHMAHAG 119

Db 61 QAPARNSDLTTHGGGNGADVGQSDSSIDLTOGFGNSATLDPWNGKSHMTYKQFG 120

Qy 120 GNNAAALVNTASDSSVWVROVGFNNATANQY 151

Db 121 GNGAAVDDTASNTSTVNTVQVGFNNATANQY 152

RESULT 5

H85665

hypothetical protein csgA [imported] - Escherichia coli (strain O157:H7, substrain EDL933)

C/Species: Escherichia coli

C/Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001

C/Accession: H85665

R/Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew

Iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,

Nature 409, 529-533, 2001

A/Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A/Reference number: AB5480; PMID:21074935; PMID:11206551

A/Accession: H85665

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-152 <STO>

A/Cross-references: GB:AE005174; NID:g12514574; PIDN:AA655788.1; GSPDB:GN00145; UMGF:216

A/Experimental source: strain O157:H7, substrain EDL933

C/Genetics:

A/Genes: csgA

Query Match 65.4%; Score 502.5; DB 2; Length 152;

Best Local Similarity 68.4%; Pred. No. 1.6e-34;

Matches 104; Conservative 18; Mismatches 29; Indels 1; Gaps 1;

Qy 1 MKLLKAAFAAIIVGSGALAGVPPW-GGGGNNNGSSGPDSTLSIYOGSANAAL 59

Db 1 MKLLKAAFAAIIVGSGALAGVPPW-GGGGNNNGSSGPDSTLSIYOGSANAAL 60

Qy 60 QSDARKSETTITGSGYNGADVGQADNSTIELTONGFNNATYDQVTRVTHEMAHAG 119

Db 61 QADARNSDLTTHGGGNGADVGQSDSSIDLTOGFGNSATLDPWNGKSHMTYKQFG 120

Qy 120 GNNAAALVNTASDSSVWVROVGFNNATANQY 151

Db 121 GNGAAVDDTASNTSTVNTVQVGFNNATANQY 152

RESULT 6

curlin nucleator protein csgB precursor - Escherichia coli (strain K-12)

N/Alternate names: csgB protein; curlin nucleation component; minor curlin protein

C/Species: Escherichia coli

C/Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 01-Mar-2002

C/Accession: S70787; F64846

R/Hammer, M.; Arngvist, A.; Bian, Z.; Olsen, A.; Normark, S.

Mol. Microbiol. 18, 661-670, 1995

A/Title: Expression of two csg operons is required for production of fibronectin- and C

A/Reference number: S70783; PMID:96414468; PMID:8817489

A/Accession: S70787

A/Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-151 <HAM>

A/Cross-references: EMBL:X90754; NID:g1147558; PIDN:CAA62281.1; PID:g1147563

C/Species: Escherichia coli

A/Experimental source: strain K12, substrain WJ10

A/Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995

R/Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co

Science 277, 1453-1462, 1997

A/Title: The complete genome sequence of Escherichia coli K-12.

A/Reference number: A64720; PMID:97426617; PMID:9278503

A/Accession: F64846

A/Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-151 <BLAT>

A/Cross-references: GB:AE000205; GB:U00096; NID:g1787265; PIDN:AA674125.1; PID:g1787278

A/Experimental source: strain K-12, substrain MG1655

C/Genetics:

A/Genes: csgB

A/Map position: 23.15

C/Function:

A/Description: minor component of wild-type curli; interaction between CsgA and CsgB tr

A/Note: curli are thin, coiled fibers expressed on the surface of Escherichia coli that

and H-kininogen; in the absence of CsgA, CsgB can self-assemble into polymers

F/1-21/Domain: signal sequence #status predicted <STG>

F/22-151/Product: minor curlin chain #status predicted <MAT>

Query Match 13.6%; Score 104.5; DB 2; Length 151;

Best Local Similarity 29.1%; Pred. No. 0.087; Length 151; Gaps 3;

Matches 34; Conservative 15; Mismatches 57; Indels 11; Gaps 3;

Qy 38 SSGPDSTLSIYOGSANAALALQSPARKSETTITGSGYNGADVGQADNSTIELTONGF 97

Db 21 AAGYDLANSEYNF---AVNELSKSSFNQALITGAGTNSAQLRQGSGLAVVAQEGS 76

Qy 98 RNNATYDQVTRVTHEMAH---AGNNAAALVNTASDSSVWVROVGFNNATANQY 151

Db 77 SNRAKIDQ---TGYNILAVYIDQASANDASISQAYGNTAMITQSGSNKANITQY 129

RESULT 7

curlin subunit precursor CsgB [imported] - Escherichia coli (strain O157:H7, sub

C/Species: Escherichia coli

C/Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001

C/Accession: G90806

R/Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G

gawaara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.

DNA Res. 8, 11-22, 2001

A/Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen

A/Reference number: A98629; PMID:21156231; PMID:11258796

A/Accession: G90806

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-151 <HAY>

A/Cross-references: GB:BA000007; PIDN:BA834842.1; PID:g13360879; GSPDB:GN00154

A/Experimental source: strain O157:H7, substrain RIMD 0509952

C/Genetics:

A/Genes: Ec81419

Query Match 13.6%; Score 104.5; DB 2; Length 151;

Best Local Similarity 29.1%; Pred. No. 0.087; Length 151; Gaps 3;

Matches 34; Conservative 15; Mismatches 57; Indels 11; Gaps 3;

Qy 38 SSGPDSTLSIYOGSANAALALQSPARKSETTITGSGYNGADVGQADNSTIELTONGF 97

Db 21 AAGYDLANSEYNF---AVNELSKSSFNQALITGAGTNSAQLRQGSGLAVVAQEGS 76

Qy 98 RNNATYDQVTRVTHEMAH---AGNNAAALVNTASDSSVWVROVGFNNATANQY 151

Db 77 SNRAKIDQ---TGYNILAVYIDQASANDASISQAYGNTAMITQSGSNKANITQY 129

RESULT 8

curlin minor chain precursor, CsgA homolog [imported] - Escherichia coli (strain O157:H

C/Species: Escherichia coli

C/Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001

C/Accession: G85665

R/Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhe

Iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,

Nature 409, 529-533, 2001

A/Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A/Reference number: AB5480; PMID:21074935; PMID:11206551

A/Accession: G85665

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-151 <STO>

A/Cross-references: GB:AE005174; NID:g12514573; PIDN:AA655787.1; GSPDB:GN00145; UMGF:216

A:Experimental source: strain 0157:H7, substrain EDL933
C:Genetics:
A:Gene: csGB

Query Match 13.6%; Score 104.5; DB 2; Length 151;
Best Local Similarity 29.1%; Pred. No. 0.087;
Matches 34; Conservative 15; Mismatches 57; Indels 11; Gaps 3;

QY 38 SSGPDSLTSLIYGYGSANALALQSDARKSETTITGSGYGADVGQADNSTIELTONGF 97
DB 21 AACGYDLANSEYFN-----ANVELSKSFNQAALIGAGTINSAQLQGGSKLAVVAQKGS 76
QY 98 RNNAITDQVLTVRVTHEMAH-----AGGNALALVNQTA SSSVWVRQVGFNGNATANQY 151
DB 77 SNNAKIDQ-----TGDNLAVIYIDQAGSANDASISQGYNTAMIIQKSGNKNANTQY 129

RESULT 9

S42136
C:Species: Tetrahymena thermophila
C:Date: 19-Mar-1997 #sequence_revision 18-Jul-1997 #text_change 07-Dec-1999
C:Accession: S42136; S42135; S03650
R:Taylor, F.M.; Martindale, D.W.
A:Reference number: 88189811, PMID:8233798
A:Accession: S42136
A:Residues: 1-1748 <TAY>

A:Cross-references: EMBL:L03710; NID:g161751; PID:g161752
R:Taylor, F.M.; Martindale, D.W.
Nucleic Acids Res. 21, 4610-4614, 1993
A:Title: Retroviral-type zinc fingers and glycine-rich repeats in a protein encoded by c
A:Reference number: S42135; MUID:94051569; PMID:8233798
A:Accession: S42135
A:Molecule type: DNA

A:Residues: 1164-1174;1179-1198;1233-1252;1285-1293;1297-1309;1316-1326;1331-1341;1343-1
A:Cross-references: EMBL:L03710
R:Martindale, D.W.; Taylor, F.M.
Nucleic Acids Res. 16, 2189-2201, 1988
A:Title: Multiple introns in a conjugation-specific gene from Tetrahymena thermophila.
A:Reference number: S03650; MUID:88189811, PMID:3357771
A:Accession: S03650
A:Molecule type: DNA

A:Residues: 236-250, 'T', 252-255, 'N', 257-773 <MAR>
A:Cross-references: EMBL:X06462
C:Genetics:
A:Gene: cnjB

A:Genetic code: SGC5
A:Introns: 85/3; 136/1; 157/3; 201/2; 290/2; 327/3; 499/1; 573/2; 607/3; 708/3; 777/3; 8
C:Keywords: zinc finger
F:1164-1450/Region: glycine-rich
F:1451-1464/Region: zinc finger CCHC motif
F:1478-1491/Region: zinc finger CCHC motif
F:1501-1514/Region: zinc finger CCHC motif
F:1530-1543/Region: zinc finger CCHC motif
F:1555-1568/Region: zinc finger CCHC motif
F:1579-1592/Region: zinc finger CCHC motif
F:1602-1615/Region: zinc finger CCHC motif
F:1626-1748/Region: glycine-rich

Query Match 13.2%; Score 101; DB 2; Length 1748;
Best Local Similarity 29.7%; Pred. No. 2.6;
Matches 38; Conservative 18; Mismatches 32; Indels 40; Gaps 7;

QY 25 QMGCGGHNHG---GNSGPDSTLSIYGSANALALQSDARKSETTITGSGYGADV 81
DB 1640 QFGGGSNGSGSGSGSDWN-----CGSNVQES-TTSSGSGWS----- 1680
QY 82 GCGADNSTIELTONGFERNATYDQVLTVRVTHEMAHAGGNALVNQTA SSSVWVRQV 141
DB 1681 -SSGND-----TGGGWSNDNQO-----QENETGGGGSNSNQTNNESS----- 1722

QY 142 FGNNATAN 149
DB 1723 WGSNNQAS 1730

RESULT 10

JC6040
C:Species: Salmonella enteritidis
C:Date: 31-Dec-1996 #sequence_revision 31-Dec-1996 #text_change 08-Oct-1999
C:Accession: JC6040
R:Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Baner, P.A.; Kay, W.W.
J. Bacteriol. 178, 662-667, 1996
A:Title: Salmonella enteritidis agfBAC operon encoding thin, aggregative fimbriae.
A:Reference number: JC6039; MUID:96146512; PMID:8550497
A:Accession: JC6040
A:Molecule type: DNA

A:Residues: 1-151 <COL>
A:Cross-references: GB:U43280; NID:g1184712; PID:AA43598.1; PID:g1184713
A:Experimental source: strain 276755-3b
C:Genetics:
A:Gene: agfB

C:Function:
A:Description: minor component of thin aggregative fimbriae
A:Note: fimbriae bind to fibronectin, plasminogen, tissue plasminogen activator
C:Keywords: fimbria
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-151/Product: fimbria protein agfB #status predicted <MAT>

Query Match 13.0%; Score 99.5; DB 2; Length 151;
Best Local Similarity 28.2%; Pred. No. 0.22;
Matches 33; Conservative 16; Mismatches 51; Indels 17; Gaps 4;

QY 51 GSANALALQSDARKSE-----TTTOSGYGNGADVGQ-GADNST-----IELTQ 94
DB 14 GAPGIATATNYDLARSEYNFAVNELSKSFNQAALIGAGTINSAQLQGGSKLAVVAQKGS 73
QY 95 NCFRNATYDQVLTVRVTHEMAHAGGNALVNQTA SSSVWVRQVGFNGNATANQY 151
DB 74 ECGNRAKVDQAGNVPAY-IEQTNANDASISQAYNSAIIQKSGNKNANTQY 129

RESULT 11

AH0635
nucleation component of curlin monomers [imported] - Salmonella enterica subsp. enteric
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002

C:Accession: AH0635
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher
th, T.; Connerton, P.; Cronin, A.; Davies, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar
, S.; Moulé, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica sero
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Status: preliminary
A:Accession: AH0635
A:Molecule type: DNA

A:Residues: 1-151 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD08267.1; PID:g16502314; GSPDB:GN00176
C:Genetics:
A:Gene: STY1180

Query Match 13.0%; Score 99.5; DB 2; Length 151;
Best Local Similarity 28.2%; Pred. No. 0.22;
Matches 33; Conservative 16; Mismatches 51; Indels 17; Gaps 4;

QY 51 GSANALALQSDARKSE-----TTTOSGYGNGADVGQ-GADNST-----IELTQ 94
DB 14 GAPGIATATNYDLARSEYNFAVNELSKSFNQAALIGAGTINSAQLQGGSKLAVVAQKGS 73
QY 95 NCFRNATYDQVLTVRVTHEMAHAGGNALVNQTA SSSVWVRQVGFNGNATANQY 151

Db 74 EGENNRAKVDQAGNYNFAY-IEQTGNANDASISQSAYGNSAAIIQKSGNKANITQY 129

RESULT 12

hypothetical glycine-rich protein [imported] - *Sinorhizobium meliloti* (strain 1021) mag
C:Species: *Sinorhizobium meliloti*
C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C:Accession: B59565
C:Author: T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A:Title: The complete sequence of the 1,683-kb *psymb* megaplaemid from the N2-fixing endo
A:Reference number: A95842; MUID:21396508; PMID:11481431
A:Accession: B59565
A:Status: preliminary
A:Molecule type: DNA
A:Reads: 1-2174 <KUR>
A:Cross-references: GB:AL591985; PID:CAC49389.1; PID:g15140875; GSPDB:GN00167
A:Experimental source: strain 1021, megaplaemid *psymb*
R:Gilbert, F.; Ffman, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kallman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
Weibull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, S.H.; Wong, K.; Yeh, K.
A:Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.
A:Reference number: A96039; MUID:21368234; PMID:11474104
A:Contents: annotation
C:Genetics:
A:Gene: SMD21548
A:Genome: plasmid

Query Match 13.0%; Score 99.5; DB 2; Length 2174;
 Best Local Similarity 23.0%; Pred. No. 4.3;
 Matches 32; Conservative 14; Mismatches 27; Gaps 2.

QY 11 AIVSGSALAGV P QWGGGNGGNGSSGPDSTLSIYQYGSANALALQSDARKSETTI 700

Db 892 AVFTKGDDSHGLAQSIGGGGNGSASS-----MSSTAGTVDTIS 932

71 TQSGYNGADVQCADNSTIETQNGFRNATYDQLVTRVTHEMAHAGNNALVNQTA 130

Db 933 ADIGSVGGSGSAGNGAVTAAKDSASVTTADYANAIVQSIGGGGNGV----- 986

QY 131 SDSSVMVRQVGFNNATAN 149

Db 987 --GSVNSKEIGSGFNLTAN 1003

RESULT 13

C:Conserved hypobacterial protein AtYu4768 [Imported] - Agrobacterium tumefaciens (strain C58)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C:Accession: AD31143
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.;
Ergase, G.; Gillet, P.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McCellen, J.;
Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E. W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AD31143
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-145 <KUR>
A:Cross-references: GB:AE006689; PIDN:RAL45562.1; PID:G17743277; GSPDB:GN00187
A:Experimental source: strain C58 (Dupont)
C:Genetic(s)
A:Gene: AtYu4768
A:Map position: linear chromosome

Query Match	12.8%;	Score 98;	DB 2;	Length 145;
Best Local Similarity	27.1%;	Pred. No. 0.29;		
Matches 38;	Conservative 25;	Mismatches 57;	Indels 20;	Gaps 5;

3 LKVAAFVAIVSGSALAGVP-----QWG-----GGGNHNGGN-----SSGPDST 44

1 MIRKSFIASALVALVGLSAAAPAMANDVRIEQYGWSNSAGGAQEGYGNRIPTYQNGGYNR 60

45 LSIYQGSANALALQSDARKSETTITQSGYNGADVQGDNSTIELTQNGFRNNATYD 104

61 IVGHQYGRHLS-AVGQEGHDNYGSTTQNGNRNVAGIGQFGSNHTTILTDGNGNIAGV 119

105 QLVTRVVTHEMAHAGNNAA 124

120 Q-VGRGCSANVSQGGNDNVA 138

RESULT 14

```

;Hypothetical protein AGR_L_228 [imported] - Agrobacterium tumefaciens (strain C58, Cere
;Species: Agrobacterium tumefaciens
;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
;Accession: H98144
;Author: Goddard, B., Hinkley, G., Gattung, S., Miller, N., Blanchard, M., Goldman, B., Goldan
;Goddard, B., Hinkley, G., Gattung, S., Miller, N., Blanchard, M., Goldman, B., Goldan
;Liu, F., Wolman, C., Allinger, M., Doughy, D., Scott, C., Lappas, C., Markelz, B.
;Science 294, 2323-2328, 2001
;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
;Reference number: A97359; M01D:21608551; PMID:11743194
;Accession: H98144
;Status: preliminary
;Molecule type: DNA
;Residues: 1-145 <KUR>
;Cross-references: GB:AE007870; PIDN:AAK68682.1; PID:g15158413; GSPDB:GN00170
;Genetics:
;Gene: AGR_L_228
;Map position: linear chromosome

```

Query Match	12.8%	Score 98;	DB 2;	Length 145;
Best Local Similarity	27.1%	Pred. No. 0.29;		
Matches	38;	Conservative	25;	Mismatches 57;
			Indels	20;
			Gaps	5

3 LTKVAFAIVSGSALAGVP-----QWG---GGGNHNGGN---SSGPDST 44

1 MIRKSFIALVALVGLSAAAPAMANDVRIEQYGWSNSAGGAQEGYGNRIRTYQNNGYNR 60

45 LSIYQYGSANALALQSDARKSETTITQSGYNGADVQGDNSTIELTQNGFRNNATYD 104

61 IVGHQYGRHNL-S-AVGQEGHDNYGSTTQNGNRNVAGIGQFGSNHTTILTDQNGNIACV 119

105 QLVTRVVTHEMAHAGGNAA 124

120 Q-VGRGCSANTVSGGNDNVA 138

RESULT 15

Probable PPE protein - *Mycobacterium tuberculosis* (strain H37RV)
 Species: *Mycobacterium tuberculosis*
 Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 15-Sep-2003
 Accession: F70825
 Cole, S.T.; Davies, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.T.; Birtles, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Ainsdreadam, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome
 Reference number: A70500; M01ID:98295987; PMID:9634230
 Accession: F70825
 Status: preliminary; nucleic acid sequence not shown; translation not shown
 Molecule type: DNA
 Residues: 1645 <COL>
 Cross-references: GB:AL021958; GB:AL123456; NID:g3261536; PIDN:CAA17522.1; PID:e125322

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM proteoln - protein search, using sw model

Run on: March 11, 2004, 18:14:48 ; Search time 6.3 Seconds
(without alignments)
1248.031 Million cell updates/sec

Title: US-09-543-407-30

Perfect score: 768
Sequence: 1 MKLLKVAAPAIIVSGSALA.....DSSVMROYFGNNATANOY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	681	88.7	151	1 CSQA_SALTY	P55225 salmonella
2	517	67.3	151	1 CSQA_SALTY	P28307 escherichia
3	502.5	65.4	152	1 CSQA_SALTY	P39828 escherichia
4	104.5	13.6	151	1 CSQA_SALTY	P55226 salmonella
5	99.5	13.0	151	1 CSQA_SALTY	P55226 salmonella
6	99.5	13.0	151	1 CSQA_SALTY	P55226 salmonella
7	92.5	12.0	1028	1 OVO_DROME	P51521 drosophila
8	91.5	11.9	1327	1 TNK1_HUMAN	O95271 homo sapien
9	90	11.7	1656	1 OMPB_RICUA	O06653 r outer mem
10	89	11.6	959	1 N100_YEAST	O02629 saccharomyc
11	88	11.5	1778	1 N189_SCHPO	O90624 schizosacch
12	86.5	11.3	262	1 VG38_BPT2	P07875 bacterioph
13	86.5	11.3	590	1 GP63_LRIDO	P23223 leishmania
14	86	11.2	646	1 GP63_LTIME	P43150 leishmania
15	85.5	11.1	1567	1 ICEN_XANCT	P18127 xanthomonas
16	85	11.1	599	1 GP63_LEIMA	P15706 leishmania
17	85	11.1	850	1 PRSN_PIERA	O08148 leishmania
18	85	11.1	1655	1 OMPB_RICCN	O90864 pteris rapa
19	85	11.1	1655	1 OMPB_RICCN	O90864 pteris rapa
20	84.5	11.0	678	1 YF48_MYCTU	Q10778 mycobacteri
21	84.5	11.0	760	1 YB1L_ECOLI	P75780 escherichia
22	84.5	11.0	1196	1 ICEN_PESBX	O93479 pseudomonas
23	84.5	11.0	1469	1 ICEN_PESBX	O93479 pseudomonas
24	84	10.9	363	1 OMPB_KLEPN	O48473 klebsiella
25	84	10.9	576	1 DEAF_DROME	O24180 drosophila
26	83.5	10.8	1148	1 ICEN_PESBX	O30611 pseudomonas
27	83	10.8	401	1 YK03_CAEEL	P34291 caenorhabdi
28	82.5	10.7	592	1 CEA_CITFR	P04480 citrobacter
29	81.5	10.6	172	1 CH18_DROME	P07184 drosophila
30	81.5	10.6	347	1 MS2A_PLAF2	O03646 plasmodium
31	81.5	10.6	392	1 HME1_HUMAN	O05925 homo sapien
32	81	10.5	165	1 GRP1_ORISA	P25074 oryza sativ
33	81	10.5	1845	1 Z236_HUMAN	Q9U136 homo sapien

ALIGNMENTS

RESULT 1	CSQA_SALTY	STANDARD	PRT	151 AA
AC	P55225			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Major curlin subunit precursor (Fimbriae SEF17).			
GN	CSQA OR AGFA OR STM1144 OR STY1181 OR T1776.			
OS	Salmonella typhimurium			
OS	Salmonella enteritidis			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OC	Enterobacteriaceae; Salmonella.			
OX	NCHI_TaxID=602, 601, 592;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=S. typhimurium; STRAIN=SR-11;			
RX	MEDLINE=98117058; PubMed=9457880;			
RA	Romling U., Bian Z., Hammar M., Sierralta W.D., Normark S.;			
RT	"Curli fibers are highly conserved between Salmonella typhimurium and			
RT	Escherichia coli with respect to operon structure and regulation.";			
RL	J. Bacteriol. 180:722-731(1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=S. typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;			
RX	MEDLINE=21534948; PubMed=11677609;			
RA	McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,			
RA	Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,			
RA	Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,			
RA	Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,			
RA	Waterston R., Wilson R.K.;			
RT	"Complete genome sequence of Salmonella enterica serovar Typhimurium			
RT	LT2.";			
RL	Nature 413:852-856(2001).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=S. typhimurium; STRAIN=CT18;			
RX	MEDLINE=21534947; PubMed=11677608;			
RA	Parthill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,			
RA	Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,			
RA	Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,			
RA	Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,			
RA	Fetellin T., Hamlin N., Haque A., Hien T.T., Holtroyd S., Jagels K.,			
RA	Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,			
RA	Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,			
RA	Whitehead S., Barrett B.G.;			
RT	"Complete genome sequence of a multiple drug resistant Salmonella			
RT	enterica serovar Typhimurium CT18.";			
RL	Nature 413:848-852(2001).			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=S. typhimurium; STRAIN=Ty2 / ATCC 700931;			
RX	MEDLINE=22531367; PubMed=12644504;			
RA	Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,			
RA	Burland V., Kodoyanni V., Schwartz D.C., Blattner F.R.;			

34	80.5	10.5	534	1	GPB2_YEAST	O01722 saccharomyc
35	80.5	10.5	874	1	ALB6_AZOVI	O92702 azotobacter
36	80.5	10.5	880	1	GUN4_THERFU	P26221 thermomonos
37	80.5	10.5	1210	1	ICEN_PSEFL	P09615 pseudomonas
38	80.5	10.5	1258	1	ICEN_ERHRE	P16239 erythrina her
39	80	10.4	380	1	PLVA_COLGL	O00374 colletoctic
40	80	10.4	491	1	YK98_MYCTU	O10707 mycobacteri
41	80	10.4	850	1	PRSN_FIBER	O99736 pteris bras
42	80	10.4	1571	1	C3G_DROME	O77086 drosophila
43	80	10.4	5560	1	SPEN_DROME	O88x83 drosophila
44	79.5	10.4	163	1	HCV_NATPH	P39442 natronomona
45	79.5	10.4	663	1	DUS8_MOUSE	O09112 mus musculu

RT "Comparative genomics of *Salmonella enterica* serovar Typhl strains Ty2
RT and CT18.";
RT J. Bacteriol. 185:2330-2337(2003).

RP SEQUENCE FROM N.A.
RC SPECIES=S. enteritidis; STRAIN=27655-3B;
RX MEDLINE=96146512; PubMed=8550497;
RA Collinson S.K., Clouthier S.C., Doran J.L., Baner P.A., Kay W.W.;
RT "Salmonella enteritidis agfBAC operon encoding thin, aggregative
RT fimbriae.";
RT J. Bacteriol. 178:662-667(1996).

RP SEQUENCE OF 21-151 FROM N.A.
RC SPECIES=S. enteritidis; STRAIN=27655-3B;
RX MEDLINE=94013373; PubMed=8104955;
RA Doran J.L., Collinson S.K., Burian J., Santos G., Todd E.C.D.,
RA Munro C.K., Kay C.M., Baner P.A., Peterkin P.I., Kay W.W.;
RT "DNA-based diagnostic tests for *Salmonella* species targeting agfA,
RT the structural gene for thin, aggregative fimbriae.";
RT J. Clin. Microbiol. 31:2263-2273(1993).

RP SEQUENCE OF 21-33.
RC SPECIES=S. enteritidis; STRAIN=27655-3B;
RX MEDLINE=91310586; PubMed=1677357;
RA Collinson S.K., Emeody L., Mueller K.-M., Trust T.J., Kay W.W.;
RT "Purification and characterization of thin, aggregative fimbriae from
RT *Salmonella enteritidis*.";
RT J. Bacteriol. 173:4773-4781(1991).
-1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN.

CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC
CC THE SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

DR EMBL: AJ002301; CAA05317.1;
DR EMBL: AE008749; AAL20074.1;
DR EMBL: AL627268; CAD20074.1;
DR EMBL: AE016840; AAO69399.1;
DR EMBL: U43280; AAC43599.1;
DR PIR: JC6039; JC6039.
DR StyGene: SG10608; CSGA.
KW Fimbria; Signal; Complete proteome.
FT SIGNAL 1 20
FT CHAIN 21 151
FT CONFLICT 134 151
FT MAJOR CURLIN SUBUNIT.
FT SVMWVGFQGNATANY -> DSYTQVAS (IN
FT REF. 6).
SQ SEQUENCE 151 AA; 15305 MW; B7DAC0D16B621359 CRC64;

Query Match 88.7%; Score 681; DB 1; Length 151;
Beet Local Similarity 90.7%; Pred. No. 1.8e-50;
Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

QY 1 MKLKVAAFAAIVVSGSALAGVVPWGGGNGGNGGSSGPDSTLSTIYQGSNNAALAQ 60
DB 1 MKLKVAAFAAIVVSGSALAGVVPWGGGNGGNGGSSGPDSTLSTIYQGSNNAALAQ 60
QY 61 SDARKSETTTTQSGYGADVGAGADNSTLTETQNGFRNNAATDQLVTVVTHEMAHAG 120
DB 61 SDARKSETTTTQSGYGADVGAGADNSTLTETQNGFRNNAATDQLVTVVTHEMAHAG 120
QY 121 NNAALVWQTASDSSVWVWVGFQGNATANY 151
DB 121 NNAALVWQTASDSSVWVWVGFQGNATANY 151

RESULT 2

CSGA_ECOLI STANDARD; PRT; 151 AA.
ID CSGA_ECOLI
AC P28307;
DT 01-DEC-1992 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Major curlin subunit precursor.
GN CSGA OR B1042.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OC NCBI_TaxID=562;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / W3110.
RX MEDLINE=93211294; PubMed=8459772;
RA Olsen A., Argyvist A.;
RT "The Rpos sigma factor relieves H-NS-mediated transcriptional
RT repression of *csga*, the subunit gene of fibronectin-binding curli in
RT *Escherichia coli*.";
RT Mol. Microbiol. 7:523-536(1993).

RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MC4100.
RX MEDLINE=96414468; PubMed=8817489;
RA Hammer M., Argyvist A., Bian Z., Olsen A., Normark S.;
RT "Expression of two *csa* operons is required for production of
RT fibronectin- and Congo red-binding curli polymers in *Escherichia coli*
RT K-12.";
RT Mol. Microbiol. 18:661-670(1995).

RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of *Escherichia coli* K-12.";
RT Science 277:1233-1238(1997).

RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajinaka M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sampaio G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RT "A 718-kb DNA sequence of the *Escherichia coli* K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map.";
RT DNA Res. 3:137-155(1996).

RP SEQUENCE OF 21-40.
RC STRAIN=K12 / YMEI;
RX MEDLINE=93023873; PubMed=1357528;
RA Argyvist A., Olsen A., Pfeiffer J., Russell D.G., Normark S.;
RT "The Ctl protein activates cryptic genes for curli formation and
RT fibronectin binding in *Escherichia coli* HB101.";
RT Mol. Microbiol. 6:2443-2452(1992).

RP SEQUENCE OF 21-31.
RC MEDLINE=91310586; PubMed=1677357;
RA Collinson S.K., Emeody L., Trust T.J., Kay W.W.;
RT "Purification and characterization of thin, aggregative fimbriae from
RT *Salmonella enteritidis*.";
RT J. Bacteriol. 173:4773-4781(1991).
-1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN.


```

CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L04979; AAA2616.1; -
CC DR EMBL: X90754; CA62282.1; -
CC DR EMBL: AE000205; AAC7426.1; -
CC DR EMBL: D90741; BAA35832.1; -
CC DR EMBL: D90742; BAA35840.1; -
CC DR PIR: S70788; S70788.
CC DR Ecogen: EG11489; CSGA.
CC Fimbrtia; Signal; Complete proteome.
CC FT SIGNAL 1 20
CC FT CHAIN 21 151 MAJOR CURLIN SUBUNIT.
CC FT CONFLICT 7 7 A -> B (IN REF. 1).
CC FT SEQUENCE 151 AA; 15049 MW; C003470D208D395F CRC64;
CC SQ
CC
CC Query Match 67.3%; Score 517; DB 1; Length 151;
CC Best Local Similarity 69.5%; Pred. No. 9.7e-37; Indels 0; Gaps 0;
CC Matches 105; Conservative 17; Mismatches 29;
CC
CC QY 1 MLLKVAAPAAIIVSGSALAGVPPQWGGGNNHGGSGPDSLSTIYQGSANAALALQ 60
CC DB 1 MLLKVAAPAAIIVSGSALAGVPPQWGGGNNHGGSGNSPSELINITYQGGNSALALQ 60
CC
CC QY 61 SDARKSETTITQSGYGADVGQGANSTIELTQNGFRNNATYDQLVTVVTHEMAHAG 120
CC DB 61 TDARNSDLITTOHGGGADVGQGSDDSIDLTQGFNGSATLDDQNGKSHMTVKQFG 120
CC
CC QY 121 NNAAVNOTASDSSVMRWQVFGNNATANOY 151
CC DB 121 GNGAAVDDTASNSTVTVQVFGNNATANOY 151
CC
CC RESULT 3
CC CSGA_ECO57 STANDARD; PRT; 152 AA.
CC ID CSGA_ECO57 STANDARD; PRT; 152 AA.
CC AC Q93024;
CC DT 28-FEB-2003 (Rel. 41, Created)
CC DT 28-FEB-2003 (Rel. 41, Last sequence update)
CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
CC DE Major curlin subunit precursor.
CC GN CSGA OR Z1676 OR ECS1420.
CC OS Escherichia coli O157:H7.
CC OC Bacteriella; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
CC OC Enterobacteriaceae; Escherichia.
CC NCBI_TaxID=83334;
CC
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN=O157:H7 / ATCC 43895;
CC RX MEDLINE=21218556; PubMed=11319125;
CC RA Unlich G.A., Keen J.B., Elder R.O.;
CC RT "Mutations in the csgD promoter associated with variations in curli
CC expression in certain strains of Escherichia coli O157:H7."
CC RL Appl. Environ. Microbiol. 67:2367-2370(2001).
CC
CC RN [2]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
CC RX MEDLINE=21074935; PubMed=1126551;
CC RA Berna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
CC Rose D.U., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
CC Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
CC Grodzick E.J., Davis N.W., Lim A., Dimailanta E.T., Potamoudis K.,
CC Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
CC Welch R.A., Blattner F.R.;
CC RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.",
CC Nature 409:529-533(2001).

```

```

RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohnubo E., Nakayama K., Murata T., Tanaka M., Toke T.,
RA Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shingawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12."
RL DNA Res. 8:11-22(2001).
CC
CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN.
CC
CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF275733; AAK53212.1; -
CC DR EMBL: AE005315; AAG55788.1; -
CC DR EMBL: AP002554; BAB34843.1; -
CC DR PIR: D90806; D90806.
CC DR PIR: H85665; H85665.
CC Fimbrtia; Signal; Complete proteome.
CC FT SIGNAL 1 20
CC FT CHAIN 21 152 MAJOR CURLIN SUBUNIT.
CC FT SEQUENCE 152 AA; 15099 MW; EE2D2D94DDE91243 CRC64;
CC
CC Query Match 65.4%; Score 502.5; DB 1; Length 152;
CC Best Local Similarity 68.4%; Pred. No. 1.6e-35; Indels 1; Gaps 1;
CC Matches 104; Conservative 18; Mismatches 29;
CC
CC QY 1 MLLKVAAPAAIIVSGSALAGVPPQWGGGNNHGGSGPDSLSTIYQGSANAALALQ 59
CC DB 1 MLLKVAAPAAIIVSGSALAGVPPQWGGGNNHGGSGNSPSELINITYQGGNSALALQ 60
CC
CC QY 60 QSDARKSETTITQSGYGADVGQGANSTIELTQNGFRNNATYDQLVTVVTHEMAHAG 119
CC DB 61 QDARNSDLITTOHGGGADVGQGSDDSIDLTQGFNGSATLDDQNGKSHMTVKQFG 120
CC
CC QY 120 GNNAAVNOTASDSSVMRWQVFGNNATANOY 151
CC DB 121 GNGAAVDDTASNSTVTVQVFGNNATANOY 152
CC
CC RESULT 4
CC CSGB_ECOLI STANDARD; PRT; 151 AA.
CC ID CSGB_ECOLI STANDARD; PRT; 151 AA.
CC AC P39828;
CC DT 01-FEB-1995 (Rel. 31, Created)
CC DT 01-OCT-1996 (Rel. 34, Last sequence update)
CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
CC DE Minor curlin subunit precursor.
CC GN CSGB OR B1041 OR Z1675 OR ECS1419.
CC OS Escherichia coli, and
CC OS Escherichia coli O157:H7.
CC OC Bacteriella; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
CC OC Enterobacteriaceae; Escherichia.
CC NCBI_TaxID=562; 83334;
CC
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN=K12 / MC4100;
CC RX MEDLINE=96414468; PubMed=8817489;
CC RA Hammar M., Arngvist A., Bian Z., Olsen A., Normark S.;
CC RT "Expression of two csg operons is required for production of
CC fibronectin- and congo red-binding curli polymers in Escherichia coli

```

RT K-12." ;
 RL Mol. Microbiol. 18:661-670 (1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of *Escherichia coli* K-12." ;
 RL Science 277:1233-1474 (1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97061202; PubMed=8905232;
 RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
 RA Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
 RA Sampaio G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horuchi T.;
 RT "A 718-kb DNA sequence of the *Escherichia coli* K-12 genome
 corresponding to the 12.7-28.0 min region on the linkage map." ;
 RL DNA Res. 3:137-155 (1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Posfai G., Hackett J., Link S., Boutin A., Shao Y., Miller L.,
 RA Godbeck E.J., Davis N.W., Lim A., Dimalanta R.T., Potamousis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7." ;
 RL Nature 409:529-533 (2001).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsuda E., Nakayama K., Murata T., Tanaka M., Toke T.,
 RA Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,
 RA Kubura S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohaemorrhagic *Escherichia coli*
 O157:H7 and genomic comparison with a laboratory strain K-12." ;
 RL DNA Res. 8:11-22 (2001).
 RN [6]
 RP SEQUENCE OF 1-21 FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=95157246; PubMed=7854117;
 RA Arngvist A., Olsen A., Normark S.;
 RT "Sigma S-dependent growth-phase induction of the *csga* promoter in
Escherichia coli can be achieved in vivo by sigma 70 in the absence
 of the nucleoid-associated protein H-NS." ;
 RL Mol. Microbiol. 13:1021-1032 (1994).
 CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
 CC CURLIN MONOMERS.
 CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL, X90754; CAA62281.1; -.

DR EMBL; AE000205; AAC74125.1; -.
 DR EMBL; D90741; BAA35831.1; -.
 DR EMBL; AE005315; AAC55787.1; -.
 DR EMBL; AP002554; BAB34842.1; -.
 DR PIR; C90806; C90806.
 DR PIR; G85665; G85665.
 DR PIR; S70787; S70787.
 DR EcoGene; EG12621; csgrB.
 KW Rimbria; Signal; Complete proteome.
 FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 22 151 MINOR CURLIN SUBUNIT.
 SQ SEQUENCE 151 AA; 15882 MW; B18D26B964014B8 CRC64;
 Query Match 13.6%; Score 104.5; DB 1; Length 151;
 Best Local Similarity 29.1%; Pred. No. 0.033;
 Matches 34; Conservative 15; Mismatches 57; Indels 11; Gaps 3;
 QY 38 SSGPDSLTSLIYGYGSANALALQSDARKSETTITGSGYGNAGDVGQAGDNSTIELTONGF 97
 DB 21 AAGYDLANSBYN---AVNELSKSEFNQAAITGGGTNNNSQLRGSGSKLAVVAQEGS 76
 QY 98 RNNATYDQLVTRVTHEM--AGGNALALVNGTSSDSSVWRQYGFNNATANY 151
 DB 77 SNRAKIDQ---TGDYNLAVIDQAGSANDASISOGAYGNTAMITOKSGSNKANTQY 129
 RESULT 5
 CSGB_SALTI STANDARD; PRT; 151 AA.
 AC Q827M3;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Minor curlin subunit precursor.
 GN CSGB OR STY1180 OR T1177.
 OS Salmonella typhi.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 OX NCBI_Taxid=601;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CT18;
 RX MEDLINE=21534947; PubMed=11677608;
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,
 RA Baker S., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
 RA Cronin A., Davis P., Brooks K., Chillingworth T., Connor P.,
 RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
 RA Krogan A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
 RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
 RA Whitehead S., Barrrell B.G.;
 RT "Complete genome sequence of a multiple drug resistant *Salmonella*
 RT enterica serovar Typhi CT18." ;
 RL Nature 413:848-852 (2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Ty2 / ATCC 700931;
 RX MEDLINE=22531367; PubMed=12644504;
 RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
 RA Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;
 RT "Comparative genomics of *Salmonella enterica* serovar Typhi strains Ty2
 RT and CT18." ;
 RL J. Bacteriol. 185:2330-2337 (2003).
 CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
 CC CURLIN MONOMERS.
 CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

CC used by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch.)
 CC -----
 DR EMBL: AL627269; CAD08267.1; -
 DR EMBL: AE016840; AAO69400.1; -
 KW Fimbricia; Signal; Complete proteome.
 FT CHAIN 1 21 POTENTIAL.
 FT CHAIN 22 151 MINOR CURLIN SUBUNIT.
 SQ SEQUENCE 151 AA; 16254 MW; 161C54326E573495 CRC64;
 Query Match 13.0%; Score 99.5; DB 1; Length 151;
 Best Local Similarity 28.2%; Pred. No. 0.088;
 Matches 33; Conservative 16; Mismatches 51; Indels 17; Gaps 4;
 QY 51 GSANALALQSDARKSE-----TTTOSGYGAGDVGQ--GADNST-----IELTQ 94
 DB 14 GAGGATATNTYDLARSEYNFAVNELSKSSFNQALIIQVGTDSARVRQEGSKLSVISQ 73
 QY 95 NCFRNATYDQVTRVVTHEMAHAGGNALVNQTASDSSVWVROYVGFNNATANY 151
 DB 74 EGGNRAKVDQAGNVPAY-IEOTGNANDASISQAYGNSAIIQKSGSKANITQY 129
 RESULT 6
 CSGB_SALTY STANDARD; PRT; 151 AA.
 AC P55236;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Minor curlin subunit precursor (Fimbrin SEF17 minor subunit).
 OS CSGB OR AGFA OR STM1143.
 OS Salmonella typhimurium, and
 OS Salmonella enteritidis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 NCBI_TaxID=602, 592;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S.typhimurium, STRAIN=SR-11;
 RA MEDLINE=98117058; PubMed=9457880;
 RA Romling U., Bian Z., Hammar M., Sierra M.D., Normark S.;
 RT "Curli fibers are highly conserved between Salmonella typhimurium and
 RT Becherichia coli with respect to operon structure and regulation.";
 RL J. Bacteriol. 180:722-731(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S.typhimurium, STRAIN=LT2 / SGSC1412 / ATCC 700720;
 RA MEDLINE=21534948; PubMed=11677609;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulyanay E.,
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
 RT LT2";
 RL Nature 413:852-856(2001).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S.enteritidis, STRAIN=27655-3B;
 RA MEDLINE=9616512; PubMed=8550497;
 RA Collinson S.K., Cloutier S.C., Doran J.L., Baner P.A., Kay W.M.;
 RT "Salmonella enteritidis agfBC operon encoding thin, aggregative
 RT fimbriae";
 RL J. Bacteriol. 178:662-667(1996).
 CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
 CC CURLIN MONOMERS.
 CC -1- SIMILARITY: BELONGS TO THE CSGB/CSGB FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch.)
 CC -----
 DR EMBL: AJ002301; CA05316.1; -
 DR EMBL: AE008749; AAL20073.1; -
 DR EMBL: UA3280; AAC43598.1; -
 DR PIR: JC6040; JC6040.
 DR StyGene; SG10609; CSGB.
 KW Fimbricia; Signal; Complete proteome.
 FT CHAIN 1 21 POTENTIAL.
 FT CHAIN 22 151 MINOR CURLIN SUBUNIT.
 SQ SEQUENCE 151 AA; 16182 MW; C0FC5430E6DD361D CRC64;
 Query Match 13.0%; Score 99.5; DB 1; Length 151;
 Best Local Similarity 28.2%; Pred. No. 0.088;
 Matches 33; Conservative 16; Mismatches 51; Indels 17; Gaps 4;
 QY 51 GSANALALQSDARKSE-----TTTOSGYGAGDVGQ--GADNST-----IELTQ 94
 DB 14 GAGGATATNTYDLARSEYNFAVNELSKSSFNQALIIQVGTDSARVRQEGSKLSVISQ 73
 QY 95 NCFRNATYDQVTRVVTHEMAHAGGNALVNQTASDSSVWVROYVGFNNATANY 151
 DB 74 EGGNRAKVDQAGNVPAY-IEOTGNANDASISQAYGNSAIIQKSGSKANITQY 129
 RESULT 7
 CSGB_OVO_DROME STANDARD; PRT; 1028 AA.
 AC P51521; O9XZU4;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE OVO protein (Shaven baby protein).
 GN OVO OR SVB.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Ovary;
 RA MEDLINE=95021209; PubMed=7935398;
 RA Garfinkel M.D., Wang J., Liang Y., Mahowald A.P.;
 RT "Multiple products from the shavenbaby-ovo gene region of Drosophila
 RT melanogaster: relationship to genetic complexity.";
 RL Mol. Cell. Biol. 14:6809-6818(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Oregon-R;
 RX MEDLINE=91293102; PubMed=1712294;
 RA Mevel-Ninio M.T.M., Terracol R., Kafatos F.C.;
 RT "The ovo gene of Drosophila encodes a zinc finger protein required
 RT for female germ line development";
 RL EMBO J. 10:2259-2266(1991).
 CC -1- FUNCTION: REQUIRED FOR SURVIVAL AND DIFFERENTIATION OF FEMALE GERM
 CC LINE CELLS. PLAYS A ROLE IN GERM LINE SEX DETERMINATION.
 CC -1- SUBCELLULAR LOCATION: Nuclear (potential).
 CC -1- DEVELOPMENTAL STAGE: FIRST APPEARS IN THE GERMARIIUM AND
 CC ACCUMULATES IN NURSE CELLS DURING OOGENESIS. STORED IN THE EGG,
 CC BUT IS RAPIDLY LOST IN THE EMBRYOS EXCEPT FOR ITS CONTINUED
 CC PRESENCE IN THE GERM LINE PRECURSOR POLE CELLS.
 CC -1- SIMILARITY: Contains 4 C2H2-type zinc fingers.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U11383; AAB60216.1; -
 DR EMBL; X59772; CAB36921.1; ALT_SEQ.
 DR PIR; A56038; A56038.
 DR HSSP; P07248; 2ADR.
 DR TRANSFAC; T00669; -
 DR FlyBase; FBgn0003028; oyo.
 DR InterPro; IPR007087; Znf_C2H2.
 DR Pfam; PF00096; zf-C2H2; 3.
 DR SMART; SM00355; Znf_C2H2; 4.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
 DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 3.
 DR Zinc-finger; Metal-binding; DNA-binding; Repeat; Nuclear protein;
 KW Transcription regulation.
 FT DOMAIN 62 66 POLY-ALA.
 FT DOMAIN 72 77 POLY-GLY.
 FT DOMAIN 80 85 POLY-GLY.
 FT DOMAIN 98 108 POLY-GLY.
 FT DOMAIN 144 152 POLY-HIS.
 FT DOMAIN 153 159 POLY-ASN.
 FT DOMAIN 336 339 POLY-GLN.
 FT DOMAIN 347 353 POLY-GLN.
 FT DOMAIN 357 361 POLY-GLN.
 FT DOMAIN 410 414 POLY-GLN.
 FT DOMAIN 418 422 POLY-GLN.
 FT DOMAIN 426 432 POLY-GLN.
 FT DOMAIN 445 453 POLY-GLN.
 FT DOMAIN 456 459 POLY-GLN.
 FT DOMAIN 466 474 POLY-GLN.
 FT DOMAIN 497 517 POLY-ALA.
 FT DOMAIN 524 529 POLY-SER.
 FT DOMAIN 549 558 POLY-ALA.
 FT DOMAIN 639 651 POLY-ALA.
 FT DOMAIN 717 725 POLY-ALA.
 FT DOMAIN 797 802 POLY-GLN.
 FT DOMAIN 820 823 POLY-GLN.
 FT DOMAIN 874 886 POLY-GLN.
 FT ZN_FING 874 896 C2H2-TYPE 1.
 FT ZN_FING 902 924 C2H2-TYPE 2.
 FT ZN_FING 930 953 C2H2-TYPE 3.
 FT ZN_FING 969 992 C2H2-TYPE 4.
 FT CONFLICT 647 647 A -> R (IN REF. 2).
 SQ SEQUENCE 1028 AA; 110620 MW; D7068BB2BC0FEF77 CRC64;

Query Match 12.0%; Score 92.5; DB 1; Length 1028;
 Best Local Similarity 26.0%; Pred. No. 2.9;
 Matches 38; Conservative 15; Mismatches 66; Indels 27; Gaps 4;

QY 3 LLLKVAFAIVVSSALAGVVPQWGGGNNHGGSSGPDSTLSTIYGSANMALALQSD 62
 DB 59 LQNNAAAAYMSASSG-----GGCTGAGGGAGSGPGGGGANSGGGGGG----- 104
 QY 63 ARKSETTTTOSGYNGADVGGAGDNSTELTONGPRNNATYDQVTRVVTHEMAHAGNN 122
 DB 105 -----GGNGYINCAGVG-GPNNSLDGNLILNFAVSVMYNSKFNHHHHHHHNNNN 155
 QY 123 AALVNGTASDSVWVRQVGRNNMTA 148
 DB 156 -----NNNNGGQTSWGHGHPFYGGNPSA 177

RESULT 8
 TNK1 HUMAN
 ID TNK1_HUMAN STANDARD; PRT; 1327 AA.
 AC 095271; 095272;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Tankyrase 1 (EC 2.4.2.30) (TNK1) (Tankyrase I) (TNKS-1) (TRF1-
 DE interacting ankyrin-related ADP-ribose polymerase).
 GN TNK1 OR TNK51 OR TINI1 OR TINI1 OR PARP.
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 CC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
 RC TISSUE=Testis;
 RX MEDLINE=99040105; PubMed=9822378;
 RA Smith S., Giriat L., Schmitt A., de Lange T.;
 RT "Tankyrase, a poly(ADP-ribose) polymerase at human telomeres.";
 RL Science 282:1484-1487(1998).
 RN [2]
 RP SUBCELLULAR LOCATION.
 RX MEDLINE=99454782; PubMed=10523501;
 RA Smith S., de Lange T.;
 RT "Cell cycle dependent localization of the telomeric PARP, tankyrase,
 RT to nuclear pore complexes and centrosomes.";
 RL J. Cell Sci. 112:3649-3656 (1999).
 RN [3]
 RP FUNCTION, AND PHOSPHORYLATION.
 RX MEDLINE=20556282; PubMed=10988299;
 RA Chi N.-W., Lodish H.F.;
 RT "Tankyrase is a Golgi-associated mitogen-activated protein kinase
 RT substrate that interacts with IRAP in GLUT4 vesicles.";
 RL J. Biol. Chem. 275:38437-38444(2000).
 RN [4]
 RP FUNCTION, AND MUTAGENESIS OF HIS-1184 AND GLU-1291.
 RX MEDLINE=21602874; PubMed=11739745;
 RA Cook B.D., Dynek J.N., Chang W., Shostak G., Smith S.;
 RT "Role for the related poly(ADP-ribose) polymerases tankyrase 1 and 2
 RT at human telomeres.";
 RL Mol. Cell. Biol. 22:332-342(2002).
 CC -1- FUNCTION: May regulate vesicle trafficking and modulate the
 CC subcellular distribution of SLC24A/Glut4-vesicles. Has PARP
 CC activity and can modify TRF1, and thereby contribute to the
 CC regulation of telomere length.
 CC -1- CATALYTIC ACTIVITY: NAD(+) + {ADP-D-ribose} (N)-acceptor =
 CC nicotinamide + (ADP-D-riboseyl) (N+1)-acceptor.
 CC -1- SUBUNIT: Oligomerizes and associates with TNKS2. Interacts with
 CC the cytoplasmic domain of LAMP2/Octase in SLC24A/Glut4-vesicles.
 CC Binds to the N-terminus of telomeric TRF1 via the ANK repeats.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic, associated with the Golgi and
 CC with juxtanuclear SLC24A/Glut4-vesicles. A minor proportion is
 CC also found at nuclear pore complexes and around the pericentriolar
 CC matrix of mitotic centrosomes. During interphase, a small fraction
 CC of TNKS is found in the nucleus, associated with TRF1.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC Name=2;
 CC IsoId=095271-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=095271-2; Sequence=VSP_004538; VSP_004539;
 CC Note=No experimental confirmation available;
 CC -1- TISSUE SPECIFICITY: Ubiquitous; highest levels in testis.
 CC -1- PTM: Upon insulin-stimulation, phosphorylated on serine residues
 CC by MAPK kinases.
 CC -1- PTM: ADP-ribosylated (-auto).
 CC -1- SIMILARITY: Belongs to the PARP family.
 CC -1- SIMILARITY: Contains 15 ANK repeats.
 CC -1- SIMILARITY: Contains 1 sterile alpha motif (SAM) domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF082556; AAC79841.1; -

DR EMBL; AF082557; AAC79842.1; -
 DR EMBL; AF082558; AAC79843.1; -
 DR EMBL; AF082559; AAC79844.1; -
 DR HSSP; Q00420; IAMC.
 DR Genew; HGNC:11941; TNKS.
 DR MM; 603303; -
 DR GO; GO:0000781; C:chromosome, telomeric region; IDA.
 DR GO; GO:0003950; P:NAD ADP-ribosyltransferase activity; IDA.
 DR GO; GO:0005515; F:protein binding; IPI.
 DR GO; GO:0007004; P:telomerase-dependent telomere maintenance; NAS.
 DR InterPro; IPR002110; ANK.
 DR InterPro; IPR001660; SAM.
 DR Pfam; PF000023; ank; 19.
 DR Pfam; PF00536; SAM; 1.
 DR PRINTS; PRO1415; ANKYRIN.
 DR SMART; SM00248; ANK; 17.
 DR SMART; SM00454; SAM; 1.
 DR PROSITE; PSS0088; ANK_REPEAT; 15.
 DR PROSITE; PSS0297; ANK_REPEAT; 1.
 DR PROSITE; PSS0105; SAM_DOMAIN; 1.
 DR TRANSFERASE; Glycosyltransferase; NAD; Golgi stack; Telomere;
 KW Nuclear protein; Repeat; ANK repeat; ADP-ribosylation;
 KM Phosphorylation; Alternative splicing.
 FT REPEAT 215 247 ANK 1.
 FT REPEAT 248 280 ANK 2.
 FT REPEAT 281 313 ANK 3.
 FT REPEAT 368 400 ANK 4.
 FT REPEAT 401 433 ANK 5.
 FT REPEAT 434 466 ANK 6.
 FT REPEAT 521 555 ANK 7.
 FT REPEAT 557 589 ANK 8.
 FT REPEAT 590 622 ANK 9.
 FT REPEAT 683 715 ANK 10.
 FT REPEAT 716 748 ANK 11.
 FT REPEAT 749 781 ANK 12.
 FT REPEAT 836 868 ANK 13.
 FT REPEAT 869 901 ANK 14.
 FT REPEAT 902 934 ANK 15.
 FT DOMAIN 1030 1089 SAM.
 FT DOMAIN 1176 1327 PAMP.
 FT DOMAIN 9 14 POLY-HS.
 FT DOMAIN 27 34 POLY-PRO.
 FT DOMAIN 128 134 POLY-SER.
 FT DOMAIN 137 145 POLY-SER.
 FT VARSPLIC 641 643 EST -> GHS (in isoform 2).
 FT VARSPLIC 641 643 /FTRID-VSP 004538.
 FT VARSPLIC 644 1327 Missing (in isoform 2).
 FT MUTAGEN 1184 1184 /FTRID-VSP 004539.
 FT MUTAGEN 1291 1291 H->A: LOSS OF ACTIVITY; WHEN ASSOCIATED WITH A-1184.
 FT MUTAGEN 1291 1291 E->A: LOSS OF ACTIVITY; WHEN ASSOCIATED WITH A-1184.
 FT SEQUENCE 1327 AA; 142010 MW; E14DE985C710B957 CRC64;

Query Match 11.9%; Score 91.5; DB 1; Length 1327;
 Best Local Similarity 29.4%; Pred. No. 4.7;
 Matches 42; Conservative 17; Mismatches 65; Indels 19; Gaps 6;

QY 6 VAAPAAI-VVSGSALGAVPQWGGGNNHNGGNSGSDSTLSIYQVGSANAAALALQSDAR 64
 DB 99 VAAAPVPAVSTSSAAGVAPNPAAGSGNNSSPSSSTSS-SSSSPSSPSSSIAESPEAA 157
 QY 65 KSETTIT-OSGNGADVDGCGADNSTIELTONGFRNNATVDLVTRVTHEMAHAGG 120
 DB 158 GVGSTAPLPGGAGPGGVPVAVSGALRELLACRND-----VSRV--KRLVDAN 206
 QY 121 NNAALVNOTASDSSVMROVFG 143
 DB 207 VNAK--DMAGRKSSPLHFAFG 227

RESULT 9
 OMPB_RICUA

ID OMPB_RICUA STANDARD; PRT; 1656 AA.
 AC 006653;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Outer membrane protein B precursor (168 kDa surface-layer protein)
 DE (Surface protein antigen) (cell surface antigen 5) (Scs5) (rmpB)
 DE (rmp B) [contains: 120 kDa surface-exposed protein (Surface protein antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide].
 GN OMPB.
 OS Rickettsia japonica.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 OC NCBI_Taxid=35790;
 RN 11;
 RP SEQUENCE FROM N.A.
 RC STRAIN=YH;
 RA Uchiyama T.;
 RT "Sequencing of the gene encoding the protein rmp B of Rickettsia japonica";
 RL Submitted (May-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (By similarity).
 CC -1- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-layer with hexagonal symmetry.
 CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMP FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
 CC
 CC
 CC EMBL; AB003681; BAA20138.1; -
 DR InterPro; IPR006315; Autotransport.
 DR InterPro; IPR005546; Autotransporter.
 DR Pfam; PF03797; Autotransporter; 1.
 DR TIGRPFam; TIGR01414; autotransp_bar1; 2.
 FT CHAIN 1 1338 120 kDa SURFACE-EXPOSED PROTEIN.
 FT CHAIN 1339 1656 32 kDa BETA PEPTIDE.
 FT DOMAIN 528 533 POLY-GLY.
 FT SEQUENCE 1656 AA; 168097 MW; 3132A69C9D05999F CRC64;

Query Match 11.7%; Score 90; DB 1; Length 1656;
 Best Local Similarity 23.5%; Pred. No. 8;
 Matches 40; Conservative 21; Mismatches 57; Indels 52; Gaps 6;

QY 6 VAAPAAI-VVSGSALGAVPQWGGGNNHNGGNSGSDSTLSIYQVGSANAAALALQSDAR 65
 DB 509 VLAAGAITLDGSAIT-----TGDIHGNGG-----GALQSLTLANDAIK 547
 QY 66 SETTTTOSG---YNGADVDGCGADNSTIELTONGFRNNATVDLVTRVTHEMAHAGG 104
 DB 548 ---TLTLAGANITISAGGTINPOANGTITKLITSTONNIYVDCDLAIATDQTGVVDSLT 604
 QY 105 QLVTRVTHEMAHAGGNNAL-----VNOTASDSSVMROVFGNNAT 147
 DB 605 NQOTLITSTIGIIGANNITTLQGFNIGSKITLNGENVAINELVIGNNS 654

RESULT 10
 ID N100_YEAST STANDARD; PRT; 959 AA.
 AC 002629;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)

DT 15-MAR-2004 (Rel. 43, last annotation update)
 DE Nucleoporin NUP100/NSP100 (Nuclear pore protein NUP100/NSP100).
 GN NUP100 OR NSP100 OR YKL068W OR YKL336.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93054906; PubMed=1385442;
 RA Wente S.R., Rount M.P., Blobel G.;
 RL "A new family of yeast nuclear pore complex proteins.";
 J. Cell Biol. 119:705-723 (1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288c;
 RX MEDLINE=94378724; PubMed=8091863;
 RA Rasmussen S.W.;
 RT "Sequence of a 20.7 kb region of yeast chromosome XI includes the
 RT NUP100 gene, an open reading frame (ORF) possibly representing a
 RT nucleoside diphosphate kinase gene, tRNAs for His, Val and Trp in
 RT addition to seven ORFs with weak or no significant similarity to
 RT known proteins.";
 RL Yeast 10:569-574 (1994).
 CC -1- FUNCTION: Essential component of nuclear pore complex.
 CC Nucleoproteins may be involved in both binding and translocation of
 CC the proteins during nucleocytoplasmic transport.
 CC -1- SUBCELLULAR LOCATION: Nuclear pore complex.
 CC -1- DOMAIN: Contains G-L-F-G repeats.
 CC -1- SIMILARITY: Belongs to the GLFG family of nucleoporins.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; 215035; CAA57875.1; -;
 DR EMBL; X75780; CAA53406.1; -;
 DR EMBL; 228068; CAA81905.1; -;
 DR PIR; B44402; B44402. -;
 DR Germonline; 139824; -;
 DR SGD; S0001551; NUP100.
 DR GO; GO:0005643; C:nuclear pore; IDA.
 DR InterPro; IPR007230; Nucleoporin.
 DR InterPro; IPR004325; Nucleoporin_FG.
 DR Pfam; PF04096; Nucleoporin; 1. -;
 DR Pfam; PF03093; Nucleoporin_FG; 25.
 DR KW Nuclear protein; Transport; Repeat.
 FT DOMAIN 33 571
 FT 29 X 6 AA APPROXIMATE REPEATS OF
 FT G-L-F-G.
 SQ SEQUENCE 959 AA; 99988 MW; D3985FF901BBA51 CRC64;
 Query Match 11.6%; Score 89; DB 1; Length 959;
 Best Local Similarity 25.9%; Pred. No. 5.3;
 Matches 35; Conservative 20; Mismatches 52; Indels 28; Gaps 6;
 QY 27 GGGGNHNGGNSGSDSTLSIY--QYGSANAALALQSDARKSETTTQSGYGN-----GA 79
 DB 183 GNGSNIIFGAGNNSGNTTGLFGNQSSAFCTNNQSSLFQGSQNTNNAFGNQNLGGS 242
 QY 80 DVGQ-----GADSTIELTONGFRNATYDQLVTVVTHMAHAGGANAALVNQTA 130
 DB 243 SFGSKPVGSGSLFQGSNNTLTGNTN--NRNGLFQGM-----NSSNGSSNGGLFGQNS 293
 QY 131 SDSSVWVROYGFGNN 145
 DB 294 MNSST---QGVFGON 305
 RESULT 11

N189_SCHPO STANDARD; PRT; 1778 AA.
 ID N189_SCHPO
 AC Q9UTK4; P78796;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, last sequence update)
 DT 28-FEB-2003 (Rel. 41, last annotation update)
 DE Nucleoporin nup189 (Nuclear pore protein nup189).
 GN NUP189 OR SPAC1486.05.
 OS Schizosaccharomyces pombe (fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 CC Schizosaccharomycetes.
 NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitch E.,
 RA Rutherford J., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Shelton D., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
 RA Welfens L., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Meestl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
 RA Beger P., Zimmermann W., Wedler H., Wambutt R., Purrelle B.,
 RA Gottlieb A., Cadieu E., Dreano S., Gloux S., Lelaude V., Motier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucabas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Rhode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cernetti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakovski G.V., Useery D., Barrell B.G., Nurse P.;
 RT "The genome sequence of Schizosaccharomyces pombe.";
 RL Nature 415:871-880 (2002).
 RN [2]
 RP SEQUENCE OF 1428-1778 FROM N.A.
 RC STRAIN=PR745;
 RX MEDLINE=98162722; PubMed=9501991;
 RA Yoshida S., Kato K., Nakai K., Okayama H., Nojima H.;
 RT "Identification of open reading frames in Schizosaccharomyces pombe
 RT cDNAs.";
 RL DNA Res. 4:363-369 (1997).
 RN [3]
 RP IDENTIFICATION, AND SUBUNIT.
 RX MEDLINE=22263608; PubMed=12376568;
 RA Tange Y., Hirata A., Nwa O.;
 RT "An evolutionarily conserved fission yeast protein, Nedi, implicated
 RT in normal nuclear morphology and chromosome stability, interacts with
 RT Dps, Plm1/RCL1 and an essential nucleoporin.";
 RL J. Cell Sci. 115:4375-4385 (2002).
 CC -1- SUBUNIT: Monomer. Interacts with ned1.
 CC -1- SUBCELLULAR LOCATION: Nuclear pore complex (by similarity).
 CC -1- DOMAIN: Contains G-L-F-G repeats.
 CC -1- SIMILARITY: Belongs to the GLFG family of nucleoporins.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL; AL133357; CAB62415.1; -;

DR EMBL; D89145; BAI1807.1; -
 DR PIR; T50074; T50074.
 DR HSSP; P22629; 1SMF.
 DR GenDB; Spombe; SPAC1486.05; -
 DR GO; GO:0016020; C:membrane; ISS.
 DR GO; GO:0005643; C:nuclear pore; ISS.
 DR GO; GO:0006606; P:protein-nucleus import; ISS.
 DR InterPro; IPR007230; Nucleoporin2.
 DR InterPro; IPR004325; Nucleoporin2.
 DR Pfam; PF04096; Nucleoporin2; 1.
 DR Pfam; PF03093; Nucleoporin FG; 24.
 KM Nuclear protein; Transport; Repeat.
 FT DOMAIN 9 625 GLY-RICH.
 FT DOMAIN 233 596 THR-RICH.
 FT DOMAIN 393 442 ASN-RICH.
 SQ SEQUENCE 1778 AA; 189575 MW; 1D2AED5D7927AD6C CRC64;

Query Match 11.5%; Score 88; DB 1; Length 1778;
 Best Local Similarity 27.3%; Pred. No. 13;
 Matches 38; Conservative 16; Mismatches 65; Indels 20; Gaps 5;

QY 15 SGSLAGVVPQWGGGNGHNGGNSGCPDSTL-----SIYOGSANALALQSPDKSETT 69
 DB 400 SGTMGTCU---RFGANNNTAANTATPTSTFGNNSNPFSGANNNAATPSGFGSTT 456
 QY 70 ITQSGYNGADVQAGADNSTIELTONGFRNATYDQLVTRVTHMAHAGNNALVNOT 129
 DB 457 TTFA--SGGFSFGGNANNA-----PKPAFGSTATTAPKPACTGLFGGLAGANTNTATNAT 510
 QY 130 ASDSSVWVRQVFGNNATA 148
 DB 511 GTGGSL-----FGNANTA 523

RESULT 12
 Vg38_BPT2 STANDARD; PRT; 262 AA.
 AC P07875;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-AUG-1988 (Rel. 08, Last annotation update)
 DT 01-AUG-1990 (Rel. 15, Last annotation update)
 DE Receptor recognizing protein (Protein Gp38).
 OS Bacteriophage T2.
 CC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
 CC T4-like viruses.
 OK NCBI_TaxID=10664;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA MEDLINE=87283911; PubMed=3302276;
 RX Riede I., Drexler K., Bachbach M.L., Hemming U.;
 RT "RNA sequence of genes 38 encoding a receptor-recognizing protein of
 bacteriophages T2, K3 and of K3 host range mutants.";
 RL J. Mol. Biol. 194;31-39(1987).
 CC -1- FUNCTION: Vg38 is at the tip of the long tail fibers and serves as
 the phage recognition site for the cellular receptor.
 CC -1- MISCELLANEOUS: THIS PHAGE USES OUTER MEMBRANE PROTEINS OMPF AND TTR
 AS RECEPTORS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X05312; CAA28935.1; -
 DR PIR; S00275; S00275.
 DR InterPro; IPR007933; Tail_fibre_GP38.
 DR Pfam; PF05268; GP38; 1.
 KM Fiber protein; Phage recognition.
 SQ SEQUENCE 262 AA; 25801 MW; 0567366918FC745 CRC64;

Query Match 11.3%; Score 86.5; DB 1; Length 262;
 Best Local Similarity 44.8%; Pred. No. 2;
 Matches 26; Conservative 4; Mismatches 19; Indels 9; Gaps 3;

QY 27 GGGGNGGNSGSPSTLSIYOGSANALALQSPDKSETTITQSGYNGADVQG 84
 DB 175 GGGGPRFVGKXIGSDSL-----GSNASL---TDAGTGCTTP-QYGANGGANGANG 223

RESULT 13
 GP63_LEIDO STANDARD; PRT; 590 AA.
 AC P23223;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Leishmanolysin precursor (EC 3.4.24.36) (Cell surface protease)
 DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface
 endopeptidase).
 GN GP63.
 OS Leishmania donovani.
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 OK NCBI_TaxID=5661;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=LV9;
 RX MEDLINE=92107220; PubMed=1762629;
 RA Webb J.R., Burton L.L., McMaster R.W.;
 RT "Heterogeneity of the genes encoding the major surface glycoprotein
 of Leishmania donovani.";
 RL Mol. Biochem. Parasitol. 48:173-184(1991).
 CC -1- FUNCTION: Has an integral role during the infection of macrophages
 in the mammalian host.
 CC -1- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and
 P1' and basic residues at P2 and P3'. A model nonpeptide is
 cleaved at -Ala-Tyr-I-Leu-Lys-Lys-.
 CC -1- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
 CC -1- SIMILARITY: Belongs to peptidase family M8.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M60048; AAA29244.1; -
 DR HSSP; P08148; 1LML.
 DR MEROPS; M08.001; -
 DR InterPro; IPR006025; Pept_M_Zn_BS.
 DR InterPro; IPR001577; Peptidase_M8.
 DR Pfam; PF01457; Peptidase_M8; 1.
 DR PRINTS; PR00782; LSHMANOLYSIN.
 DR PROSITE; PS00142; ZINC_PROTEASE; 1.
 KM Hydrolyase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;
 KM Zymogen; Signal; Cell adhesion; GPI-anchor; Lipoprotein.
 FT SIGNAL 1 39
 FT PROPEP 40 87
 FT CHAIN 88 565
 FT PROPEP 566 590
 FT METAL 251 251
 FT ACT_SITE 252 252
 FT METAL 255 255
 FT METAL 321 321
 FT DISULFID 112 129
 FT DISULFID 178 217
 FT DISULFID 301 373
 FT DISULFID 380 443
 FT DISULFID 393 412
 FT DISULFID 402 477

FT DISULFID 454 498 BY SIMILARITY.
 FT DISULFID 503 553 BY SIMILARITY.
 FT DISULFID 523 546 BY SIMILARITY.
 FT CARBOHYD 287 287 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT LIPID 565 565 GPI-anchor amidated asparagine (By
 FT SEQUENCE 590 AA; 62950 MW; 0FB315D29659F58 CRC64;
 Query Match 11.3%; Score 86.5; DB 1; Length 590;
 Best Local Similarity 47.1%; Pred. No. 4.9;
 Matches 24; Conservative 4; Mismatches 18; Indels 5; Gaps 2;
 QY 103 YDQLVTVVTHMAHAGNNALVNCQASDSVM-VROVGFQ---NNATA 148
 DB 241 YDQLVTVVTHMAHAGNNALVNCQASDSVM-VROVGFQ---NNATA 148
 RESULT 14
 ID GP63_LEIME STANDARD; PRT; 646 AA.
 AC P43150;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Leishmanolysin C1 precursor (EC 3.4.24.36) (Cell surface protease)
 DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface
 DE endopeptidase).
 GN GP63-C1.
 OS Leishmania mexicana.
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 NCBI_TaxID=5665;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MNYC/B2/62/M379;
 RX MEDLINE=93149206; PubMed=8426614;
 RA Medina-Acosta E., Karses R.B., Russell D.G.,
 RT "Structurally distinct genes for the surface protease of Leishmania
 RT mexicana are developmentally regulated."
 RL Mol. Biochem. Parasitol. 57:31-46(1993).
 CC -1- FUNCTION: Has an integral role during the infection of macrophages
 CC in the mammalian host.
 CC -1- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and
 CC P1' and basic residues at P2 and P3'. A model nonapeptide is
 CC cleaved at -Ala-Tyr-|-Leu-Lys-Lys-.
 CC -1- CORACTOR: Binds 1 zinc ion per subunit (By similarity).
 CC -1- DEVELOPMENTAL STAGE: Expressed in both the promastigote and the
 CC amastigote forms.
 CC -1- SIMILARITY: Belongs to peptidase family M8.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X64394; CAA45733.1; -
 CC PIR; S19916; S19916.
 CC DR HSSP; P08148; 1LML.
 CC DR MEROPS; M08.001; -
 CC DR GLYCOSITEDB; P43150; -
 CC DR InterPro; IPR000577; Peptidase_M8.
 CC DR Pfam; PF01457; Peptidase_M8; 1.
 CC DR PRINTS; PR00782; LSHMANOLYSIN.
 CC DR PROSITE; PS00142; ZINC_PROTEASE; 1.
 CC KM Hydroxylase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;
 CC Zymogen; Signal; Cell adhesion; Multigene family.
 FT SIGNAL 1 39 POTENTIAL.
 FT PROPEP 40 102 ACTIVATION PEPTIDE (POTENTIAL).
 FT CHAIN 103 646 LEISHMANOLYSIN C1.
 FT METAL 266 266 ZINC (CATALYTIC) (BY SIMILARITY).

FT ACT_SITE 267 267 BY SIMILARITY.
 FT METAL 270 270 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 336 336 ZINC (CATALYTIC) (BY SIMILARITY).
 FT DISULFID 127 144 BY SIMILARITY.
 FT DISULFID 193 232 BY SIMILARITY.
 FT DISULFID 316 388 BY SIMILARITY.
 FT DISULFID 395 458 BY SIMILARITY.
 FT DISULFID 408 427 BY SIMILARITY.
 FT DISULFID 417 492 BY SIMILARITY.
 FT DISULFID 469 513 BY SIMILARITY.
 FT DISULFID 518 568 BY SIMILARITY.
 FT DISULFID 538 561 BY SIMILARITY.
 FT CARBOHYD 86 86 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 297 297 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 399 399 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 409 409 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 433 433 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 445 445 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 466 466 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 501 501 N-LINKED (GLCNAC. .) (POTENTIAL).
 SO SEQUENCE 646 AA; 65054 MW; FE48DDC78C1DB0A CRC64;
 Query Match 11.2%; Score 86; DB 1; Length 646;
 Best Local Similarity 94.4%; Pred. No. 6;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 103 YDQLVTVVTHMAHAGG 120
 DB 256 YDQLVTVVTHMAHAGV 273
 RESULT 15
 ID ICEN_XANCT STANDARD; PRT; 1567 AA.
 AC P18127;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Ice nucleation protein.
 GN INAX.
 OS Xanthomonas campestris (pv. translucens).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 OC Xanthomonadaceae; Xanthomonas.
 NCBI_TaxID=343;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=X56S;
 RX MEDLINE=91080859; PubMed=2259339;
 RA Zhao J., Orser C.S.,
 RT "Conserved repetition in the ice nucleation gene inax from
 RT Xanthomonas campestris pv. translucens."
 RL Mol. Gen. Genet. 223:163-166(1990).
 CC -1- FUNCTION: Ice nucleation proteins enable bacteria to nucleate
 CC crystallization in supercooled water. (By similarity).
 CC -1- SUBCELLULAR LOCATION: Outer membrane (By similarity).
 CC -1- DOMAIN: CONTAINS 153 IMPERFECT REPEATS OF THE CONSENSUS
 CC REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.
 CC -1- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE
 CC NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.
 CC -1- SIMILARITY: Belongs to the bacterial ice nucleation protein
 CC family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X52970; CAA37140.1; -
 CC HSSP; P06620; 1INA.

DR InterPro: IPR000258; Ice_nucleatn.
DR Pfam: PF00818; Ice_nucleation; 81.
DR PRINTS: PR00327; ICENUCLEATN.
DR PROSITE: PS00314; ICE_NUCLEATION; 57.
KM Ice nucleation; Repeat; Outer membrane.
SQ SEQUENCE 1567 AA; 152548 MW; CBB451D959ECAD63 CRC64;

Query Match 11.1%; Score 85.5; DB 1; Length 1567;
Best Local Similarity 26.4%; Pred. No. 18;
Matches 39; Conservative 27; Mismatches 39; Indels 43; Gaps 10;

QY 34 CGGNSSGPDSTLSIYQGSANALALQSDARKSETTITQSGYGN-----GADV----- 81
Db 1055 GSTGTAGADSTL-IAGYSTQTA-----GSDSLT-AGYSTQTARQGSDDITAGYGS 1104
QY 82 --GQADNSTIE--LTQN-GFRNNATYDQLVTRVVTHEMA-----HAGNNALV- 126
Db 1105 TGTAGADSLIAGYSTQTAQYDSNLTAGYSTQTARQGSSTAGHDSSTLA 1164
QY 127 ---NOTASDSSVMRQVGFNNATANQ 150
Db 1165 GYGSTQTAGYNSILT--TGYSTQTAQE 1190

Search completed: March 11, 2004, 18:35:00
Job time : 7.3 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:23:54 ; Search time 30.5 Seconds
(without alignments)
1562.074 Million cell updates/sec

Title: US-09-543-407-30

Perfect score: 768
Sequence: 1 MLLKVAAPAAIVSGSALA.....DSSVMVRYQVFGNNATANY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriaph:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	672	87.5	152	2	033802 salmonella
2	570.5	74.3	150	2	07X243 citrobacter
3	534	69.5	149	2	07X240 citrobacter
4	506.5	66.0	152	16	08CW63 escherichia
5	435.5	56.7	150	2	07X237 enterobacte
6	380	49.5	76	2	054069 salmonella
7	126	16.4	502	16	08RIH4 shewanella
8	122	15.9	29	2	09S3J5 escherichia
9	112	14.6	139	16	08RIH3 shewanella
10	111.5	14.5	171	16	08SJI3 bradyrhizob
11	105	13.7	130	16	08SJI4 bradyrhizob
12	104.5	13.6	151	16	07UC21 shigella fl
13	104.5	13.6	160	16	08CW64 escherichia
14	104.5	13.6	160	16	083RU7 shigella fl
15	104	13.5	157	16	08BHG0 pseudomonas
16	102.5	13.3	3659	16	0981N6 rhizobium 1

17	102	13.3	1422	16	08EFJ3 shewanella
18	101.5	13.2	531	13	08AWA8 lampetra fl
19	101	13.2	1748	5	094821 tetrahymena
20	101	13.2	7716	16	07UWZ8 rhodospirillum
21	99.5	13.0	1209	16	089CK5 bradyrhizob
22	99.5	13.0	2174	16	092JH8 rhizobium m
23	99	12.9	362	16	089D03 bradyrhizob
24	99	12.9	441	2	087327 pseudomonas
25	98.5	12.8	648	13	093397 cyprinus ca
26	98.5	12.8	324	3	08NIV1 neosporea
27	98	12.8	145	16	08U6N9 agrobacteri
28	97.5	12.7	91	2	09S3J8 escherichia
29	97.5	12.7	477	2	07X4S5 pseudomonas
30	97.5	12.7	586	16	08EXJ2 leptospira
31	97.5	12.7	546	16	07UIC5 mycobacteri
32	97.5	12.7	646	16	053818 mycobacteri
33	97	12.6	1410	16	08CMJ0 shewanella
34	96.5	12.6	191	10	07XDR3 oryza sativ
35	96.5	12.6	313	10	09S3Y2 arabidopsis
36	95.5	12.4	151	2	07X244 citrobacter
37	95.5	12.4	154	16	089J15 bradyrhizob
38	95.5	12.4	346	5	09SR55 drosophila
39	95.5	12.4	908	5	09VM71 drosophila
40	95	12.4	151	2	07X238 enterobacte
41	95	12.4	196	10	022638 zea mays (m
42	94.5	12.3	153	16	08SJI6 bradyrhizob
43	94.5	12.3	908	5	09VBC4 drosophila
44	94.5	12.3	909	5	08T4E0 drosophila
45	94.5	12.3	1713	3	08TGE1 saccharomyc

ALIGNMENTS

RESULT 1

033802 ID 033802 PRELIMINARY; PRT; 152 AA.

AC 01-JAN-1998 (TREMBLrel. 05, Created)

DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE Agfa protein (Fragment).

GN AGFA.

OS Salmonella typhimurium.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Salmonella.

OX NCBI_Taxid=602;

RN (1)

RP SEQUENCE FROM N.A.

RX MEDLINE=98053981; PubMed=9393832;

RA Sukhopolvi S.S., Lorentz R.G., Gordon J.I., Bian Z., Pfeiffer J.D.,

RA Normark S.J., Rhen M.;

RT "Expression of thin, aggregative fimbriae promotes interaction of

RT Salmonella typhimurium SR-11 with mouse small intestinal epithelial

RT cells";

RL Infect. Immun. 65:5320-5325(1997).

DR EMBL; AJ000514; CA04151.1; -

FT NON TER 152 152

SQ SEQUENCE 152 AA; 15401 MW; 9DA7DADC2364B006 CRC64;

Query Match 87.5%; Score 672; DB 2; Length 152;
Best Local Similarity 89.4%; Pred. No. 36-46;
Matches 135; Conservative 3; Mismatches 13; Indels 0; Gaps 0;

QY	1	MLLKVAAPAAIVSGSALAGVPPQWGGGNNNGSGNSGPSTLSIYGSANALAIQ	60
DB	1	MLLKVAAPAAIVSGSAVAVPPQWGGGNNNGSGNSGPSTLSIYGSANALAIQ	60
QY	61	SAKSETTITTSYGNGADVGAGADNSPTIELTONGFRNATYDOLVTRVTHEMAHAG	120
DB	61	SAKSETTITTSYGNGADVGAGADNSPTIELTONGFRNATYDOLVTRVTHEMAHAG	120
QY	121	NNALVNOTASDSVMVRYQVFGNNATANY 151	

```

Db      121 NNAALVNOTASDSSVWVRQVGFNNATPANOY 151
      |||||
RESULT 2
ID      07X243      PRELIMINARY;      PRT;      150 AA.
AC      07X243;
DT      01-OCT-2003 (TREMBLrel. 25, Created)
DT      01-OCT-2003 (TREMBLrel. 25, Last sequence update)
DT      01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE      Curlin-csgA protein.
GN      CSGA.
OS      Citrobacter sp. Fec2.
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC      Enterobacteriaceae; Citrobacter.
OX      NCBI_Taxid=213763;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=Fec2;
RA      Zogaj X., Bokranz W., Nimtz M., Romling U.;
RT      "Production of Cellulose and Curli Fimbriae by Members of the Family
RL      Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
DR      EMBL; AJ515700; CAD56672.1; -
SQ      SEQUENCE 150 AA; 15016 MW; 1D7141BD6973DC6 CRC64;

Query Match
Best local Similarity 74.3%; Score 570.5; DB 2; Length 150;
Matches 117; Conservative 14; Mismatches 19; Indels 1; Gaps 1;

Cy      1 MKLLKVAAPFAIIVSGSALAGVVPQWGGGNNHGGGNSGPDSTLSTIYOGSANAALALQ 60
      |||||
Db      1 MKLLQVAAPFAIIVSGSALAGSVPMWGGGGG-GGGSSSGESTLSTIYOGSANAALALQ 59
      |||||
Cy      61 SDARKSETTITQSGYNGADVGGADNSTIELTONGFRNNATYDQVTRVVTHEMAHAG 120
      |||||
Db      60 SDARKSDITTIHQNGGADVGGSDNSTIDLTONGFRNNATIDQWNGKSDITVSGYGG 119
      |||||
Cy      121 NNAALVNOTASDSSVWVRQVGFNNATPANOY 151
      |||||
Db      120 HNAALVNOTASDSSVTVHQVGFNNATPANOY 150
      |||||

RESULT 3
ID      07X240      PRELIMINARY;      PRT;      149 AA.
AC      07X240;
DT      01-OCT-2003 (TREMBLrel. 25, Created)
DT      01-OCT-2003 (TREMBLrel. 25, Last sequence update)
DT      01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE      Curlin-csgA protein.
GN      CSGA.
OS      Citrobacter freundii.
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC      Enterobacteriaceae; Citrobacter.
OX      NCBI_Taxid=546;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=Fec4;
RA      Zogaj X., Bokranz W., Nimtz M., Romling U.;
RT      "Production of Cellulose and Curli Fimbriae by Members of the Family
RL      Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
DR      EMBL; AJ515701; CAD56675.1; -
SQ      SEQUENCE 149 AA; 15260 MW; 946DD52017F648FD CRC64;

Query Match
Best local Similarity 69.5%; Score 534; DB 2; Length 149;
Matches 109; Conservative 16; Mismatches 24; Indels 2; Gaps 1;

Cy      1 MKLLKVAAPFAIIVSGSALAGVVPQWGGGNNHGGGNSGPDSTLSTIYOGSANAALALQ 60
      |||||

```

```

Db      1 MKLLKVAAPFAIIVSGSALAGVVPW--GGNHGGGNSNGPDSLSLTIYOGSNNANALQ 58
      |||||
Cy      61 SDARKSETTITQSGYNGADVGGADNSTIELTONGFRNNATYDQVTRVVTHEMAHAG 120
      |||||
Db      59 SDARKSDVITTIHQNGGADVGGADNSTIELTQTFQNSATIDQWNAKADISVTFQFG 118
      |||||
Cy      121 NNAALVNOTASDSSVWVRQVGFNNATPANOY 151
      |||||
Db      119 RNQALVNOTASDSSVTVLQVGFNNATPANOY 149
      |||||

RESULT 4
ID      08CW63      PRELIMINARY;      PRT;      152 AA.
AC      08CW63;
DT      01-MAR-2003 (TREMBLrel. 23, Created)
DT      01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT      01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE      Major curlin subunit precursor.
GN      CSGA OR C1306.
OS      Escherichia coli O6.
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC      Enterobacteriaceae; Escherichia.
OX      NCBI_Taxid=217992;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=O6:HI / ATCC 700928;
RA      MEDLINE=22388234; PubMed=12471157;
RA      Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA      Raske D., Buckles E.L., Lion S.-R., Boutin A., Hackett J., Stroud D.,
RA      Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA      Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT      "Extensive mosaic structure revealed by the complete genome sequence
RT      of uropathogenic Escherichia coli.";
RL      Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).
DR      EMBL; AE016759; AAN79779.1; -
KW      Complete proteome.
SQ      SEQUENCE 152 AA; 15064 MW; 3BA57F34C1240E83 CRC64;

Query Match
Best local Similarity 66.0%; Score 506.5; DB 16; Length 152;
Matches 105; Conservative 17; Mismatches 29; Indels 1; Gaps 1;

Cy      1 MKLLKVAAPFAIIVSGSALAGVVPQW-GGGGNNHGGGNSGPDSTLSTIYOGSANAALAL 59
      |||||
Db      1 MKLLKVAAPFAIIVSGSALAGVVPQYGGGNNHGGGNSGPNLSLTIYOGGSALALQ 60
      |||||
Cy      60 QSDARKSETTITQSGYNGADVGGADNSTIELTONGFRNNATYDQVTRVVTHEMAHAG 119
      |||||
Db      61 QADARNSDITTIHQGGGADVGGSDSIDLTQRFGNSATIDQWNGKSDITVTKQFG 120
      |||||
Cy      120 GNNALVNOTASDSSVWVRQVGFNNATPANOY 151
      |||||
Db      121 GNGAALVQDTASNSSVNVTVQVGFNNATPANOY 152
      |||||

RESULT 5
ID      07X237      PRELIMINARY;      PRT;      150 AA.
AC      07X237;
DT      01-OCT-2003 (TREMBLrel. 25, Created)
DT      01-OCT-2003 (TREMBLrel. 25, Last sequence update)
DT      01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE      Curlin-csgA protein.
GN      CSGA.
OS      Enterobacter sakazakii.
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC      Enterobacteriaceae; Enterobacter.
OX      NCBI_Taxid=28141;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=Fec39;
RA      Zogaj X., Bokranz W., Nimtz M., Romling U.;

```

RT "Production of Cellulose and Curli Fimbriae by Members of the Family
 Enterobacteriaceae Isolated from the Human Gastrointestinal Tract."
 RL Infect. Immun. 72:4151-4158(2003).
 DR EMBL, AJ515702; CAD56678.1; -
 SQ SEQUENCE 150 AA; 15112 MW; 5D8BBD872DF15F3 CRC64;

Query Match 56.7%; Score 435.5; DB 2; Length 150;
 Best Local Similarity 60.3%; Pred. No. 2, 2e-27;
 Matches 91; Conservative 26; Mismatches 33; Indels 1; Gaps 1;

QY 1 MLLKVAFAAIVVSGSALAGVVPWGSGGNNHGGSSGPDSTLSIYOGSANAALAQ 60
 DB 1 MKRIKVAALAAIVVSSSANAAMTNG-GGWHGHGHGGYGGPNSLTINLYOGGNSALAQ 59
 QY 61 SPARKSETTTTGGYGNGADVGGADNSTIELTONGFRNNATYDQLVTRVTHMAHAG 120
 DB 60 TDABNSVLINISQGGGNGADVGGSDSSINLTONGFRNSATLTDQNNKSQSVNVSQYGG 119
 QY 121 NNAAIVNOTASDSSVVMRVGVGFNNATNAY 151
 DB 120 LNALVDOTASNSTVNTVTOIGFNGHATAYQY 150

RESULT 6

Q54069 PRELIMINARY; PRT; 76 AA.
 ID 054069;
 AC 054069;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE SEPI fimbria (Fragment).
 GN AGFA.
 OS Salmonella enteritidis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 NCBI_TaxID=592;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=SE30;
 RA Cox J.M., Eglezos S., Woolcock J.B.;
 RT "Virulence of Salmonella enteritidis in chickens correlates with
 colony morphology and expression of SEPI7 fimbriae."
 RL Submitted (Apr-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL, U53207; AAA98671.1; -
 FT NON TER 1
 FT NON TER 76
 SQ SEQUENCE 76 AA; 7704 MW; 2FD5411241A7BCB1 CRC64;

Query Match 49.5%; Score 380; DB 2; Length 76;
 Best Local Similarity 97.4%; Pred. No. 2, 7e-23;
 Matches 74; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 30 GNNNGCGSSGPTSTIYOGSANAALAQSPARKSETTTTGGYGNGADVGGADNST 89
 DB 1 GNNXGCGSSGPTSTIYOGSANAALAQSPARKSETTTTGGYGNGADVGGADNST 60
 QY 90 IELTONGFRNNATYDQ 105
 DB 61 IELTONGFRNNATIDQ 76

RESULT 7

Q8EIH4 PRELIMINARY; PRT; 502 AA.
 ID 08EIH4;
 AC 08EIH4;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Conserved hypothetical protein.
 GN S00865.
 OS Shewanella oneidensis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
 OC Alteromonadaceae; Shewanella.

OX NCBI_TaxID=70863;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=MR-1;
 RX MEDLINE=22297686; PubMed=12368813;
 RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
 RA Meyer T., Tsapin A., Scott J., Beaman M., Brinkac L., Daugherty S.,
 RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
 RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
 RA Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,
 RA Mueller J., Khouri H., Gill J., Uitterback T.R., McDonald L.A.,
 RA Feldlyum T.V., Smith H.O., Venter J.C., Neeson K.H., Fraser C.M.;
 RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
 Shewanella oneidensis."
 RL Nat. Biotechnol. 20:1118-1123(2002).
 DR EMBL, AE015532; JAM53941.1; -
 DR TIGR, S00865; -
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 502 AA; 52441 MW; D08CA23D6C46B2D CRC64;

Query Match 16.4%; Score 126; DB 16; Length 502;
 Best Local Similarity 27.0%; Pred. No. 0.048;
 Matches 41; Conservative 22; Mismatches 57; Indels 32; Gaps 5;

QY 29 GNNHNG-----GGN-----SSGPDSTLSIYOGSANA--ALAIQS 61
 DB 231 GNNHNGFYVALAGSENDISMEQGSNNFAYLSMTTGDDTVDITDGDGSDNTVGDLSIADI 290
 QY 62 DAKSETTTTGGYGNGADVGGADNSTIELTONGFRNNATYDQLVTRVTHMAHAGN 121
 DB 291 QGDNDITIKQGDSSGAEFGYVWGSDNDVLKORDANPATGAYGTD--NDPDLSSKGN 349
 QY 122 NNAAIVNOTASDSSVVMRVGVGFNNATNAY 149
 DB 350 NELVAPATGEDNSIRISQBDANFAYVDATGN 381

RESULT 8

Q9S3J5 PRELIMINARY; PRT; 29 AA.
 ID 09S3J5;
 AC 09S3J5;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
 DE Curlin subunit monomer (Fragment).
 GN CSGA.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 NCBI_TaxID=562;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TRANSPOSON=Insertion sequence ISI;
 RX MEDLINE=99314153; PubMed=10386375;
 RA La Ragione R.M., Collighan R.J., Woodward M.J.;
 RT "Non-curtiation of Escherichia coli O78:K80 isolates associated with
 RT ISI inserti on in csGB and reduced persistence in poultry infection."
 RL FEMS Microbiol. Lett. 175:247-253(1999).
 DR EMBL, AJ131756; CAB45380.1; -
 FT NON TER 29
 FT NON TER 29
 SQ SEQUENCE 29 AA; 2789 MW; E290DFC07ABBB243 CRC64;

Query Match 15.9%; Score 122; DB 2; Length 29;
 Best Local Similarity 89.7%; Pred. No. 0.0035;
 Matches 26; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MLLKVAFAAIVVSGSALAGVVPWGSGG 29
 DB 1 MLLKVAALAAIVVSGSALAGVVPYGGG 29

RESULT 9

```

O8EIH3
ID O8EIH3 PRELIMINARY; PRT; 139 AA.
AC O8EIH3;
DT 01-MAR-2003 (TREMblrel. 23, Created)
DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)
DE 01-MAR-2003 (TREMblrel. 23, Last annotation update)
GN S00866.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.
OX NCBI_TaxID=70863;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MR-1;
RX MEDLINE=22297666; PubMed=12368813;
RA Heidelberg J.F., Paulsen I.T., Neilson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Meehe B., Clayton R.A.,
RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
RA Vamathevan J., Weidman J., Imbraim M., Lee K., Berry K., Lee C.,
RA Mueller J., Khouri H., Gill J., Uterback T.R., McDonald L.A.,
RA Feldblyum T.V., Smith H.O., Venter J.C., Neilson K.H., Fraser C.M.;
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis."
RL Nat. Biotechnol. 20:118-123 (2002).
DR EMBL; AB015532; AAN53942.1; -.
KW TIGR; S00866; -.
SQ Complete proteome.
OY SEQUENCE 139 AA; 14811 MW; 41EC1CFAT76957920 CRC64;
Db

Query Match 14.6%; Score 112; DB 16; Length 139;
Best Local Similarity 28.3%; Pred. No. 0.14;
Matches 32; Conservative 22; Mismatches 45; Indels 14; Gaps 3;

OY 39 SGPDSLTISYQGSANALALQSDARKSETTITSGYGNAGDVGGADNSTIELTONGFR 98
Db 41 SGRDLDLIDVOOGTANOGIVQSGSDNS-AVYTOAGNDNISLVTQIGTNNEVQLLYGAQ 99

OY 99 NNATYDQLVTRVTHHEMAAGNNAALVNOTASDSSVWVRQVGNATANOY 151
Db 100 NKASITQI-----GNDILVOLNQLGS-GNFSITQIADGAISTITQY 139

RESULT 10
O89J13 PRELIMINARY; PRT; 171 AA.
AC O89J13;
DT 01-JUN-2003 (TREMblrel. 24, Created)
DT 01-JUN-2003 (TREMblrel. 24, Last sequence update)
DE 01-JUN-2003 (TREMblrel. 24, Last annotation update)
GN CsgA protein.
OS CsgA OR BL15300.
OC Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RX MEDLINE=22484998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA Sasamoto S., Watanabe A., Ideasa K., Iritschi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimo S., Tsurutoka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110."
DL DNA Res. 9:189-197(2002).
DR EMBL; AP005954; BAC50585.1; -.
KW Complete proteome.
SQ SEQUENCE 171 AA; 17448 MW; 995DB08C01498381 CRC64;

```

```

Query Match 14.5%; Score 111.5; DB 16; Length 171;
Best Local Similarity 24.8%; Pred. No. 0.19;
Matches 40; Conservative 24; Mismatches 54; Indels 43; Gaps 5;

OY 1 MKLTKVAFATVVGSAALAGVVPQWGGGNNHGGNSGPDSTLSIYQGSANALALQSDA 60
Db 40 MKLTFPASFVAVLSSAQAQA-----NTSTTVQGLVNGSSVTQ 78

OY 61 SDARKSETTITSGYGNAGDVGGAD-----NSTIELTONGFRNNATYDQLVTRVTHHEMA 116
Db 79 NGLTNDSSSTTQIGLNGASTMQGTSSPSLNNVSTVNGQGVQNSATTGV----- 128

OY 117 HAGNNNAALVNOTA-----SDSSVWVRQVGC-NNATANO 150
Db 129 -AFGNNGSATONSFGPPALQNNASASVQLSFGINTSTVSQ 168

RESULT 11
O89J14 PRELIMINARY; PRT; 130 AA.
AC O89J14;
DT 01-JUN-2003 (TREMblrel. 24, Created)
DT 01-JUN-2003 (TREMblrel. 24, Last sequence update)
DE 01-JUN-2003 (TREMblrel. 24, Last annotation update)
GN BL15299 protein.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RX MEDLINE=22484998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA Sasamoto S., Watanabe A., Ideasa K., Iritschi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimo S., Tsurutoka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110."
DL DNA Res. 9:189-197(2002).
DR EMBL; AP005954; BAC50564.1; -.
KW Complete proteome.
SQ SEQUENCE 130 AA; 12699 MW; ACFB2D66A48D260F CRC64;

Query Match 13.7%; Score 105; DB 16; Length 130;
Best Local Similarity 24.7%; Pred. No. 0.46;
Matches 37; Conservative 27; Mismatches 60; Indels 26; Gaps 4;

OY 4 LKVAFAATVVGSAALAGVVPQWGGGNNHGGNSGPDSTLSIYQGSANALALQSDA 63
Db 1 KRITVLVATALASLTITVDQ-----AGNSA-----SVLQGTNTSSISITQGS 45

OY 64 RKSETTITSGYGNAGDVGGADNSTIELT-----QNGFRNNATYDQLVTRVTHHEMAAG 120
Db 46 TSNNTTTLQFATVATATLTQGSLLTAVTAVTQGGTTATASNALNGOV-----GG 97

OY 121 NNATVNOTASDSSVWVRQVGNATANO 150
Db 98 SSSSLIGIGANNNTAGVQGLGILNGSTTLQ 127

RESULT 12
O7UCZ1 PRELIMINARY; PRT; 151 AA.
AC O7UCZ1;
DT 01-OCT-2003 (TREMblrel. 25, Created)
DT 01-OCT-2003 (TREMblrel. 25, Last sequence update)
DE 01-OCT-2003 (TREMblrel. 25, Last annotation update)
GN Minor curlin subunit.
OS CsgB OR S1108.
OC Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

```

OC Enterobacteriaceae; Shigella.
 NCBI_TaxID=623;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=2457T / ATCC 700930 / Serotype 2a;
 RX MEDLINE=22590274; PubMed=12704152;
 RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
 R Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
 RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
 RA Schwartz D.C., Blattner F.R.;
 RT "Complete genome sequence and comparative genomics of Shigella
 flexneri serotype 2a strain 2457T.";
 RL Infect. Immun. 71:2775-2786(2003).
 DR EMBL: AE016981; AAPI6542.1; -
 SQ SEQUENCE 151 AA; 15868 MW; 5D5D266B964014A0 CRC64;
 Query Match 13.6%; Score 104.5; DB 16; Length 151;
 Best Local Similarity 29.1%; Pred. No. 0.6; Indels 11; Gaps 3;
 Matches 34; Conservative 15; Mismatches 57;
 QY 38 SSGPDSLSTIYQGSANALALQSDARKSETTITGSGYNGADVQAGADNSTIELTQNGF 97
 DB 21 AAGYDLANSEYNF---AVNELSKSFNQALIGAGTNNSAQLRQGGSKLLAVVAQEGS 76
 QY 98 RNNATYDQVTRVVTHEMAH---AGGNNAALVNTQASDSSVWYRVQGFNNATANQY 151
 DB 77 SNRAKIDQ---TGDNVLAVIDQAGSANDASISQAGYNTAMTIQKSGNKANITQY 129
 RESULT 13
 Q8CM64 PRELIMINARY; PRT; 160 AA.
 AC Q8CM64;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Minor curlin subunit precursor.
 GN CSGB OR C1305.
 OS Escherichia coli O6.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 NCBI_TaxID=217992;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
 RX MEDLINE=22388234; PubMed=12471157;
 RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
 RA Raeko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
 RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
 RT "Extensive mosaic structure revealed by the complete genome sequence
 of uropathogenic Escherichia coli";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
 DR EMBL: AE016759; AAN79778.1; -
 KW Complete proteome.
 SQ SEQUENCE 160 AA; 16963 MW; 49F68448D979B86 CRC64;
 Query Match 13.6%; Score 104.5; DB 16; Length 160;
 Best Local Similarity 29.1%; Pred. No. 0.65; Indels 11; Gaps 3;
 Matches 34; Conservative 15; Mismatches 57;
 QY 38 SSGPDSLSTIYQGSANALALQSDARKSETTITGSGYNGADVQAGADNSTIELTQNGF 97
 DB 30 AAGYDLANSEYNF---AVNELSKSFNQALIGAGTNNSAQLRQGGSKLLAVVAQEGS 85
 QY 98 RNNATYDQVTRVVTHEMAH---AGGNNAALVNTQASDSSVWYRVQGFNNATANQY 151
 DB 86 SNRAKIDQ---TGDNVLAVIDQAGSANDASISQAGYNTAMTIQKSGNKANITQY 138
 RESULT 14
 Q83R07 PRELIMINARY; PRT; 160 AA.
 ID Q83R07

AC Q83R07;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Minor curlin subunit precursor, similar to CsgA.
 GN CSGB OR SP1035.
 OS Shigella flexneri.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Shigella.
 NCBI_TaxID=623;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=301 / Serotype 2a;
 RX MEDLINE=22272406; PubMed=12384590;
 RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
 RA Sun L., Xue Y., Zhao A., Gao Y., Zhang J., Yang G., Wu H., Qu D., Dong J.,
 RA Sun L., Xue Y., Zhao A., Gao Y., Zhang J., Yang G., Wu H., Qu D., Dong J.,
 RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
 Yu J.;
 RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
 through comparison with genomes of Escherichia coli K12 and O157.";
 RL Nucleic Acids Res. 30:4432-4441(2002).
 DR EMBL: AE015131; AAN42658.1; -
 KW Complete proteome.
 SQ SEQUENCE 160 AA; 16919 MW; 50269F5268D2A32F CRC64;
 Query Match 13.6%; Score 104.5; DB 16; Length 160;
 Best Local Similarity 29.1%; Pred. No. 0.65; Indels 11; Gaps 3;
 Matches 34; Conservative 15; Mismatches 57;
 QY 38 SSGPDSLSTIYQGSANALALQSDARKSETTITGSGYNGADVQAGADNSTIELTQNGF 97
 DB 30 AAGYDLANSEYNF---AVNELSKSFNQALIGAGTNNSAQLRQGGSKLLAVVAQEGS 85
 QY 98 RNNATYDQVTRVVTHEMAH---AGGNNAALVNTQASDSSVWYRVQGFNNATANQY 151
 DB 86 SNRAKIDQ---TGDNVLAVIDQAGSANDASISQAGYNTAMTIQKSGNKANITQY 138
 RESULT 15
 Q88HG0 PRELIMINARY; PRT; 157 AA.
 AC Q88HG0;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Curli fiber surface-exposed nucleator CsgB, putative.
 GN PP3398.
 OS Pseudomonas putida (strain KT2440).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 NCBI_TaxID=160489;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=2243060; PubMed=12534463;
 RA Nelson K.E., Weinel C., Paulsen I.T., Dodson R.J., Hilbert H.,
 RA Martins dos Santos V.A.P., Fouts D.E., Gill S.R., Pop M., Holmes M.,
 RA Brinac L., Beanan M., DeBoy R.T., Daugherty S., Kolonay J.,
 RA Madupu R., Nelson W., White O., Peterson J., Khouri H., Hance I.,
 RA Chris Lee P., Holzapfel E., Scanlan D., Tran K., Moazzes A.,
 RA Utechback T., Rizzo M., Lee K., Kosack D., Moestl D., Wedler H.,
 RA Lauber J., Stjepandic D., Hoheisel J., Straetz M., Helm S.,
 RA Frazer C., Eissen J., Timmis K.N., Duesterhoeft A., Tsemmler B.,
 RA Fraser C.M.;
 RT "Complete genome sequence and comparative analysis of the
 metabolically versatile Pseudomonas putida KT2440.";
 RL Environ. Microbiol. 4:799-808(2002).
 DR EMBL: AE016786; AAN69002.1; -
 DR TIGR: PP3398; -
 DR InterPro: IPR000437; Prok_Lipoprot. S.
 DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
 KW Complete proteome.
 SQ SEQUENCE 157 AA; 16175 MW; C385E98465D2F960 CRC64;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:33:51 ; Search time 171.3 Seconds
(without alignments)
860.386 Million cell updates/sec

Title: US-09-543-407-12

Perfect score: 779

Sequence: 1 MKLLKVAAPAAIVSGSALA.....VTRVTHMAHANNATANOY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 6019581 seqs, 976053577 residues

Total number of hits satisfying chosen parameters: 6019581

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending Patents_AA_Main:*

1: /cgn2_6/prodata/2/paa/PCTUS_COMB.pep.*

2: /cgn2_6/prodata/2/paa/US06_COMB.pep.*

3: /cgn2_6/prodata/2/paa/US07_COMB.pep.*

4: /cgn2_6/prodata/2/paa/US08_COMB.pep.*

5: /cgn2_6/prodata/2/paa/US081_COMB.pep.*

6: /cgn2_6/prodata/2/paa/US082_COMB.pep.*

7: /cgn2_6/prodata/2/paa/US083_COMB.pep.*

8: /cgn2_6/prodata/2/paa/US084_COMB.pep.*

9: /cgn2_6/prodata/2/paa/US085_COMB.pep.*

10: /cgn2_6/prodata/2/paa/US086_COMB.pep.*

11: /cgn2_6/prodata/2/paa/US087_COMB.pep.*

12: /cgn2_6/prodata/2/paa/US088_COMB.pep.*

13: /cgn2_6/prodata/2/paa/US089_COMB.pep.*

14: /cgn2_6/prodata/2/paa/US090_COMB.pep.*

15: /cgn2_6/prodata/2/paa/US091_COMB.pep.*

16: /cgn2_6/prodata/2/paa/US092_COMB.pep.*

17: /cgn2_6/prodata/2/paa/US093_COMB.pep.*

18: /cgn2_6/prodata/2/paa/US094_COMB.pep.*

19: /cgn2_6/prodata/2/paa/US095_COMB.pep.*

20: /cgn2_6/prodata/2/paa/US096_COMB.pep.*

21: /cgn2_6/prodata/2/paa/US097A_COMB.pep.*

22: /cgn2_6/prodata/2/paa/US097B_COMB.pep.*

23: /cgn2_6/prodata/2/paa/US098_COMB.pep.*

24: /cgn2_6/prodata/2/paa/US099A_COMB.pep.*

25: /cgn2_6/prodata/2/paa/US099B_COMB.pep.*

26: /cgn2_6/prodata/2/paa/US100_COMB.pep.*

27: /cgn2_6/prodata/2/paa/US101_COMB.pep.*

28: /cgn2_6/prodata/2/paa/US102_COMB.pep.*

29: /cgn2_6/prodata/2/paa/US103_COMB.pep.*

30: /cgn2_6/prodata/2/paa/US104_COMB.pep.*

31: /cgn2_6/prodata/2/paa/US106_COMB.pep.*

32: /cgn2_6/prodata/2/paa/US107_COMB.pep.*

33: /cgn2_6/prodata/2/paa/US60_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Query Length	DB ID	Description
-----	-----	-----	-----	-----

1	779	100.0	151	19	US-09-543-407-12	Sequence 12, Appl
2	696	89.3	151	19	US-09-543-407-14	Sequence 14, Appl
3	695	89.2	151	19	US-09-543-407-5	Sequence 5, Appl
4	690	88.6	151	6	US-08-233-642A-57	Sequence 57, Appl
5	655	84.1	151	19	US-09-543-407-24	Sequence 24, Appl
6	614	78.8	151	19	US-09-543-407-26	Sequence 26, Appl
7	613	78.7	151	19	US-09-543-407-18	Sequence 18, Appl
8	611	78.4	151	19	US-09-543-407-20	Sequence 20, Appl
9	608	78.0	131	19	US-09-543-407-31	Sequence 31, Appl
10	605	77.7	151	19	US-09-543-407-28	Sequence 28, Appl
11	604	77.5	151	19	US-09-543-407-22	Sequence 22, Appl
12	603	77.4	151	19	US-09-543-407-30	Sequence 30, Appl
13	580	74.5	151	19	US-09-543-407-16	Sequence 16, Appl
14	560	71.9	120	6	US-08-233-642A-55	Sequence 55, Appl
15	520	66.8	151	19	US-09-543-407-7	Sequence 7, Appl
16	517	66.4	151	13	US-08-978-878-4	Sequence 4, Appl
17	517	66.4	151	21	US-09-741-873B-4	Sequence 4, Appl
18	515	66.1	151	33	US-60-352-946-2	Sequence 2, Appl
19	515	66.1	151	33	US-60-444-371-2	Sequence 2, Appl
20	481	61.7	158	16	US-09-252-691-5834	Sequence 5834, Ap
21	481	61.7	158	16	US-09-252-691C-5834	Sequence 5834, Ap
22	481	61.7	158	30	US-10-417-886-5834	Sequence 5834, Ap
23	476	61.1	109	19	US-09-543-407-34	Sequence 34, Appl
24	439	56.4	131	13	US-08-978-878-2	Sequence 2, Appl
25	439	56.4	131	21	US-09-741-873B-2	Sequence 2, Appl
26	347	44.5	68	19	US-09-543-407-37	Sequence 37, Appl
27	335	43.0	109	19	US-09-543-407-35	Sequence 35, Appl
28	247.5	31.8	70	19	US-09-543-407-32	Sequence 32, Appl
29	237	30.4	48	19	US-09-543-407-39	Sequence 39, Appl
30	100.5	12.9	151	19	US-09-543-407-6	Sequence 6, Appl
31	100.5	12.9	186	16	US-09-252-691-5833	Sequence 5833, Ap
32	100.5	12.9	186	16	US-09-252-691C-5833	Sequence 5833, Ap
33	100.5	12.9	186	30	US-10-417-886-5833	Sequence 5833, Ap
34	100	12.8	445	29	US-10-369-493-20638	Sequence 20638, A
35	100	12.8	445	33	US-60-360-039-20638	Sequence 20638, A
36	100	12.8	1249	30	US-10-455-719-358	Sequence 358, App
37	100	12.8	1249	33	US-60-385-568-357	Sequence 357, App
38	100	12.8	1249	33	US-60-446-775-358	Sequence 358, App
39	100	12.8	1249	33	US-10-179-131-9317	Sequence 9317, App
40	98.5	12.6	850	20	US-09-614-150-24084	Sequence 24084, A
41	98.5	12.6	850	20	US-09-614-150A-24084	Sequence 24084, A
42	98.5	12.6	850	33	US-60-191-637-24179	Sequence 24179, A
43	98.5	12.6	850	33	US-60-191-681-19019	Sequence 19019, A
44	98.5	12.6	1028	20	US-09-614-150-14916	Sequence 14916, A
45	98.5	12.6	1028	20	US-09-614-150A-14916	Sequence 14916, A

ALIGNMENTS

RESULT 1

US-09-543-407-12

Sequence 12, Application US/09543407

GENERAL INFORMATION:

APPLICANT: White, Aaron P.

APPLICANT: Doran, James L.

APPLICANT: Collinson, S. Karen

APPLICANT: Kay, William W.

TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES

FILE REFERENCE: 920043.406

CURRENT APPLICATION NUMBER: US/09/543,407

CURRENT FILING DATE: 2000-04-05

NUMBER OF SEQ ID NOS: 59

SOFTWARE: PatSeq for Windows Version 4.0

SEQ ID NO 12

LENGTH: 151

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA

OTHER INFORMATION: sequence containing the replacement fragment

OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.

US-09-543-407-12

Query Match 100.0%; Score 779; DB 19; Length 151;
Best Local Similarity 100.0%; Pred. No. 2.9e-74;
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLKVAAPFAIIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
DB 1 MKLLKVAAPFAIIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
QY 61 SDARKSETTITQSGYGNAGADVGGADNSTIELTONGFRNNATTIDQWAKNSDITVQYGG 120
DB 61 SDARKSETTITQSGYGNAGADVGGADNSTIELTONGFRNNATTIDQWAKNSDITVQYGG 120
QY 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151
DB 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151

RESULT 2

US-09-543-407-14

Sequence 14, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIBRINAL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543,407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 14
LENGTH: 151
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-14

Query Match 89.3%; Score 696; DB 19; Length 151;
Best Local Similarity 87.6%; Pred. No. 2e-65;
Matches 141; Conservative 0; Mismatches 0; Indels 20; Gaps 2;

QY 1 MKLLKVAAPFAIIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
DB 1 MKLLKVAAPFAIIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
QY 61 SDARKSETTITQSGYGNAGADVGGADNSTIELTONGFRNNATTIDQWAKNSDITVQYGG 120
DB 61 SDARKSETTITQSGYGNAGADVGGADNSTIELTONGFRNNATTIDQWAKNSDITVQYGG 120
QY 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151
DB 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151

RESULT 3

US-09-543-407-5

Sequence 5, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIBRINAL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543,407

CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5
LENGTH: 151
TYPE: PRT
ORGANISM: Salmonella enteritidis
US-09-543-407-5

Query Match 89.2%; Score 695; DB 19; Length 151;
Best Local Similarity 90.1%; Pred. No. 2.5e-65;
Matches 136; Conservative 4; Mismatches 11; Indels 0; Gaps 0;

QY 1 MKLLKVAAPFAIIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
DB 1 MKLLKVAAPFAIIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
QY 61 SDARKSETTITQSGYGNAGADVGGADNSTIELTONGFRNNATTIDQWAKNSDITVQYGG 120
DB 61 SDARKSETTITQSGYGNAGADVGGADNSTIELTONGFRNNATTIDQWAKNSDITVQYGG 120
QY 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151
DB 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151

RESULT 4

US-08-233-642A-57

Sequence 57, Application US/08233642A
GENERAL INFORMATION:
APPLICANT: Kay, William W.
APPLICANT: Collinson, S. Karen
APPLICANT: Clouthier, Sharon C.
APPLICANT: Doran, James L.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
NUMBER OF SEQUENCES: 58
NUMBER OF SEQUENCES: -
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed and Berry
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: U.S.A.
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233,642A
FILING DATE: 26-APR-1994
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: King, Joshua
REGISTRATION NUMBER: 35,570
REFERENCE/DOCKET NUMBER: 920043.403C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEEDANBERRY
INFORMATION FOR SEQ ID NO: 57:
SEQUENCE CHARACTERISTICS:
LENGTH: 151 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-233-642A-57

Query Match 88.6%; Score 690; DB 6; Length 151;
Best Local Similarity 89.4%; Pred. No. 8.6e-65;
Matches 135; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

```

QY      1 MLLKVAAPFAIVSGSALAGVPPQWGGGNNHNGSGSPDSTLSIYOGSANAALALQ 60
      1 MLLKVAAPFAIVSGSALAGVPPQWGGGNNHNGSGSPDSTLSIYOGSANAALALQ 60
DB      61 SPARSETTITGSGNGADVGAGDNGSTIELTONGFRNNATTIDOMNAKNSDITVGOYGG 120
      61 SPARSETTITGSGNGADVGAGDNGSTIELTONGFRNNATTIDOMNAKNSDITVGOYGG 120
QY      121 NNAALVNTDQVTRVVTHEMAHANNATANQY 151
      121 NNPALVNTASDSSVWVRQVGFGNNAATANQY 151
DB      121 NNPALVNTASDSSVWVRQVGFGNNAATANQY 151

RESULT 5
US-09-543-407-24
; Sequence 24, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-24

Query Match
Best Local Similarity 84.1%; Score 655; DB 19; Length 151;
Matches 134; Conservative 0; Mismatches 0; Indels 34; Gaps 2;

QY      1 MLLKVAAPFAIVSGSALAGVPPQWGGGNNHNGSGSPDSTLSIYOGSANAALALQ 60
      1 MLLKVAAPFAIVSGSALAGVPPQWGGGNNHNGSGSPDSTLSIYOGSANAALALQ 60
DB      61 SPARSETTITGSGNGADVGAGDNGSTIELTONGFRNNATTIDOMNAKNSDITVGOYGG 120
      61 SPARSETTITGSGNGADVGAGDNGSTIELTONGFRNNATTIDOMNAKNSDITVGOYGG 120
QY      61 SPARSETTITGSGNGADVGAGDNGSTIELTONGFRNNATTIDOMNAKNSDITVGOYGG 120
      61 SPARSETTITGSGNGADVGAGDNGSTIELTONGFRNNATTIDOMNAKNSDITVGOYGG 120
DB      61 SPARSETTITGSGNGADVGAGDNGSTIELTONGFRNNATTIDOMNAKNSDITVGOYGG 120
      61 SPARSETTITGSGNGADVGAGDNGSTIELTONGFRNNATTIDOMNAKNSDITVGOYGG 120
QY      121 NNAALVNTDQVTRVVTHEMAHANNATANQY 151
      121 NNPALVNTASDSSVWVRQVGFGNNAATANQY 151
DB      121 NNPALVNTASDSSVWVRQVGFGNNAATANQY 151

RESULT 6
US-09-543-407-26
; Sequence 26, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 151

```

```

; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-26

Query Match
Best Local Similarity 78.8%; Score 614; DB 19; Length 151;
Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;

QY      1 MLLKVAAPFAIVSGSALAGVPPQWGGGNNHNGSGSPDSTLSIYOGSANAALALQ 60
      1 MLLKVAAPFAIVSGSALAGVPPQWGGGNNHNGSGSPDSTLSIYOGSANAALALQ 60
DB      61 SPARSETTITGSGNGADVGAGDNGSTIELTONGFRNNATTIDOMNAKNSDITVGOYGG 120
      61 SPARSETTITGSGNGADVGAGDNGSTIELTONGFRNNATTIDOMNAKNSDITVGOYGG 120
QY      61 SPARSETTITGSGNGADVGAGDNGSTIELTONGFRNNATTIDOMNAKNSDITVGOYGG 120
      61 SPARSETTITGSGNGADVGAGDNGSTIELTONGFRNNATTIDOMNAKNSDITVGOYGG 120
DB      61 SPARSETTITGSGNGADVGAGDNGSTIELTONGFRNNATTIDOMNAKNSDITVGOYGG 120
      61 SPARSETTITGSGNGADVGAGDNGSTIELTONGFRNNATTIDOMNAKNSDITVGOYGG 120
QY      121 NNAALVNTDQVTRVVTHEMAHANNATANQY 151
      121 NNPALVNTASDSSVWVRQVGFGNNAATANQY 151
DB      121 NNPALVNTASDSSVWVRQVGFGNNAATANQY 151

RESULT 7
US-09-543-407-18
; Sequence 18, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-18

Query Match
Best Local Similarity 78.7%; Score 613; DB 19; Length 151;
Matches 122; Conservative 7; Mismatches 22; Indels 0; Gaps 0;

QY      1 MLLKVAAPFAIVSGSALAGVPPQWGGGNNHNGSGSPDSTLSIYOGSANAALALQ 60
      1 MLLKVAAPFAIVSGSALAGVPPQWGGGNNHNGSGSPDSTLSIYOGSANAALALQ 60
DB      61 SPARSETTITGSGNGADVGAGDNGSTIELTONGFRNNATTIDOMNAKNSDITVGOYGG 120
      61 SPARSETTITGSGNGADVGAGDNGSTIELTONGFRNNATTIDOMNAKNSDITVGOYGG 120
QY      61 SPARSETTITGSGNGADVGAGDNGSTIELTONGFRNNATTIDOMNAKNSDITVGOYGG 120
      61 SPARSETTITGSGNGADVGAGDNGSTIELTONGFRNNATTIDOMNAKNSDITVGOYGG 120
DB      61 SPARSETTITGSGNGADVGAGDNGSTIELTONGFRNNATTIDOMNAKNSDITVGOYGG 120
      61 SPARSETTITGSGNGADVGAGDNGSTIELTONGFRNNATTIDOMNAKNSDITVGOYGG 120
QY      121 NNAALVNTDQVTRVVTHEMAHANNATANQY 151
      121 NNPALVNTASDSSVWVRQVGFGNNAATANQY 151
DB      121 NNPALVNTASDSSVWVRQVGFGNNAATANQY 151

RESULT 8
US-09-543-407-20
; Sequence 20, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.

```

```

; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 151
; TYPE: PRF
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-20

```

```

Query Match
Best Local Similarity 78.4%; Score 611; DB 19; Length 151;
Matches 122; Conservative 7; Mismatches 22; Indels 0; Gaps 0;

```

```

QY 1 MLLKVAAPAAIVSGSALAGVVPQWGGGNNHNGGSSGPDSTLSIYQGSANALALQ 60
DB 1 MLLKVAAPAAIVSGSALAGVVPQWGGGNNHNGGSSGPDSTLSIYQGSANALALQ 60
QY 61 SPARKSETTITGSGYNGADVGQGDNSTIELTONGFNNATTIDQWAKNSDITVGQYG 120
DB 61 SPARKSETTITGSGYNGADVGQGDNSTIELTONGFNNATTIDQWAKNSDITVGQYG 120
QY 121 NNAALVNYDQVTRVVTHEMAHANNATANQY 151
DB 121 NNAALVNYDQVTRVVTHEMAHANNATANQY 151

```

```

RESULT 9
US-09-543-407-31

```

```

; Sequence 31, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 131
; TYPE: PRF
; ORGANISM: Salmonella enteritidis
US-09-543-407-31

```

```

Query Match
Best Local Similarity 78.0%; Score 608; DB 19; Length 131;
Matches 116; Conservative 4; Mismatches 11; Indels 0; Gaps 0;

```

```

QY 21 GVPQWGGGNNHNGGSSGPDSTLSIYQGSANALALQSDARKSETTITGSGYNGAD 80
DB 1 GVPQWGGGNNHNGGSSGPDSTLSIYQGSANALALQSDARKSETTITGSGYNGAD 60
QY 81 VGGADNSTIELTONGFNNATTIDQWAKNSDITVGQYGNNAALVNYDQVTRVVTHEM 140
DB 61 VGGADNSTIELTONGFNNATTIDQWAKNSDITVGQYGNNAALVNYDQVTRVVTHEM 120
QY 141 AHANNATANQY 151
DB 121 GFGNNATANQY 131

```

```

RESULT 10
US-09-543-407-28
; Sequence 28, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 151
; TYPE: PRF
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-28

```

```

Query Match
Best Local Similarity 77.7%; Score 605; DB 19; Length 151;
Matches 121; Conservative 5; Mismatches 25; Indels 0; Gaps 0;

```

```

QY 1 MLLKVAAPAAIVSGSALAGVVPQWGGGNNHNGGSSGPDSTLSIYQGSANALALQ 60
DB 1 MLLKVAAPAAIVSGSALAGVVPQWGGGNNHNGGSSGPDSTLSIYQGSANALALQ 60
QY 61 SPARKSETTITGSGYNGADVGQGDNSTIELTONGFNNATTIDQWAKNSDITVGQYG 120
DB 61 SPARKSETTITGSGYNGADVGQGDNSTIELTONGFNNATTIDQWAKNSDITVGQYG 120
QY 121 NNAALVNYDQVTRVVTHEMAHANNATANQY 151
DB 121 NNAALVNYDQVTRVVTHEMAHANNATANQY 151

```

```

RESULT 11
US-09-543-407-22

```

```

; Sequence 22, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 22
; LENGTH: 151
; TYPE: PRF
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-22

```

```

Query Match
Best Local Similarity 77.5%; Score 604; DB 19; Length 151;
Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;

```

```

QY 1 MLLKVAAPAAIVSGSALAGVVPQWGGGNNHNGGSSGPDSTLSIYQGSANALALQ 60

```

Db 1 MKLKVAAFAIIVSGSALAGVVPWGCGGNHNGGSSGPDSTLSIYQGSANALALQ 60
QY 61 SDARKSETTTGSGYGADVGAGADNSTIETLTONGFRNNATTIDOWNAKNSDITVGGYGG 120
Db 61 SDARKSETTTGSGYGADVGAGADNSTIETLTONGFRNNATTIDOWNAKNSDITVGGYGG 120
QY 121 NNAALVNYDQLVTRVVTHEMAHANNATANY 151
Db 121 NNAALVNYDQASDSSVMVRQVGFNNATANY 151

RESULT 12
US-09-543-407-30
Sequence 30, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 30
LENGTH: 151
TYPE: PRF
ORGANISM: Artificial Sequence

QY 1 MKLKVAAFAIIVSGSALAGVVPWGCGGNHNGGSSGPDSTLSIYQGSANALALQ 60
Db 1 MKLKVAAFAIIVSGSALAGVVPWGCGGNHNGGSSGPDSTLSIYQGSANALALQ 60
QY 61 SDARKSETTTGSGYGADVGAGADNSTIETLTONGFRNNATTIDOWNAKNSDITVGGYGG 120
Db 61 SDARKSETTTGSGYGADVGAGADNSTIETLTONGFRNNATTIDOWNAKNSDITVGGYGG 120
QY 121 NNAALVNYDQLVTRVVTHEMAHANNATANY 151
Db 121 NNAALVNYDQASDSSVMVRQVGFNNATANY 151

Query Match 77.4%; Score 603; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 1.6e-55;
Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;

US-09-543-407-16
Sequence 16, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 16
LENGTH: 151
TYPE: PRF
ORGANISM: Artificial Sequence

FEATURE:
OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afga
OTHER INFORMATION: sequence containing the replacement fragment
US-09-543-407-16

Query Match 74.5%; Score 580; DB 19; Length 151;
Best Local Similarity 80.1%; Pred. No. 4.5e-53;
Matches 121; Conservative 7; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLKVAAFAIIVSGSALAGVVPWGCGGNHNGGSSGPDSTLSIYQGSANALALQ 60
Db 1 MKLKVAAFAIIVSGSALAGVVPWGCGGNHNGGSSGPDSTLSIYQGSANALALQ 60
QY 61 SDARKSETTTGSGYGADVGAGADNSTIETLTONGFRNNATTIDOWNAKNSDITVGGYGG 120
Db 61 SDARKSETTTGSGYGADVGAGADNSTIETLTONGFRNNATTIDOWNAKNSDITVGGYGG 120
QY 121 NNAALVNYDQLVTRVVTHEMAHANNATANY 151
Db 121 NNAALVNYDQASDSSVMVRQVGFNNATANY 151

RESULT 14
US-08-233-642A-55
Sequence 55, Application US/08233642A
GENERAL INFORMATION:
APPLICANT: Kay, William W.
APPLICANT: Cloutier, Sharon C.
APPLICANT: Doran, James L.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR *SALMONELLA*-
TITLE OF INVENTION: BASED VACCINES
NUMBER OF SEQUENCES: 58
NUMBER OF SEQUENCES: -
CORRESPONDENCE ADDRESS:
ADDRESSER: Seed and Berry
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: U.S.A.
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233,642A
FILING DATE: 26-APR-1994
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: King, Joshua
REGISTRATION NUMBER: 35,570
REFERENCE/DOCKET NUMBER: 920043.403C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEEDABERRY
INFORMATION FOR SEQ ID NO: 55:
SEQUENCE CHARACTERISTICS:
LENGTH: 120 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-233-642A-55

Query Match 71.9%; Score 560; DB 6; Length 120;
Best Local Similarity 100.0%; Pred. No. 4.5e-51;
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 VVPWGCGGNHNGGSSGPDSTLSIYQGSANALALQSDARKSETTTGSGYGADV 81

Db 1 VVPQWGGGNHNGGSSGPDSTLSTIYQGSANALALQSDARKSETTTTQSGYNGADV 60

Qy 82 GGGADNSTIELTONGFRNNATIDQWNAKNSDITVGYGGNNALVN 127
 Db 61 GGGADNSTIELTONGFRNNATIDQWNAKNSDITVGYGGNNALVN 106

RESULT 15

US-09-543-407-7
 ; Sequence 7, Application US/09543407
 ; GENERAL INFORMATION:
 ; APPLICANT: White, Aaron P.
 ; APPLICANT: Doran, James L.
 ; APPLICANT: Collinson, S. Karen
 ; APPLICANT: Kay, William W.
 ; TITLE OF INVENTION: BACTERIAL FIBRIAL SYSTEM FOR
 ; FILE REFERENCE: 920043.406
 ; CURRENT APPLICATION NUMBER: US/09/543,407
 ; CURRENT FILING DATE: 2000-04-05
 ; NUMBER OF SEQ ID NOS: 59
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 7
 ; LENGTH: 151
 ; TYPE: PRT
 ; ORGANISM: Escherichia coli
 US-09-543-407-7

Query Match 66.8%; Score 520; DB 19; Length 151;
 Best Local Similarity 66.9%; Pred. No. 1.le-46;
 Matches 101; Conservative 21; Mismatches 29; Indels 0; Gaps 0;

Qy 1 MLLKVAFAFAIVVSSALAGVVPQWGGGNHNGGSSGPDSTLSTIYQGSANALALQ 60
 Db 1 MLLKVAFAFAIVVSSALAGVVPQWGGGNHNGGSSGPDSTLSTIYQGSANALALQ 60
 Qy 61 SPARKSETTTTQSGYNGADVQGGADNSTIELTONGFRNNATIDQWNAKNSDITVGYGG 120
 Db 61 TDARNSDLTITQHGNGADVQGGSDSDSIDLTQRFNSATILDQWNGKNSMTVKQPCG 120
 Qy 121 NNAALVNDQLVTRVVTHEMAHANNATANQY 151
 Db 121 GNGAAYVDQTASNSVNTQVGFNNATYAHQY 151

Search completed: March 11, 2004, 19:13:09
 Job time : 172.3 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using SW model

Run on: March 11, 2004, 18:35:05 ; Search time 5.3 Seconds
(without alignments)
376.014 Million cell updates/sec

Title: US-09-543-407-12

Perfect score: 779
Sequence: 1 MKLLKVAAPAAIVSGSALA.....VTRVTHMAHANNATANY 151

Scoring table: BLOSUM62
Gap 10.0 , Gapext 0.5

Searched: 125546 seqs, 13197846 residues

Total number of hits satisfying chosen parameters: 125546

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents_AA_New:*
1: /cgn2_6/prodata/1/paa/PCT_NEW_COMB.pep:*
2: /cgn2_6/prodata/1/paa/US06_NEW_COMB.pep:*
3: /cgn2_6/prodata/1/paa/US07_NEW_COMB.pep:*
4: /cgn2_6/prodata/1/paa/US08_NEW_COMB.pep:*
5: /cgn2_6/prodata/1/paa/US10_NEW_COMB.pep:*
6: /cgn2_6/prodata/1/paa/US10_NEW_COMB.pep:*
7: /cgn2_6/prodata/1/paa/US60_NEW_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	517	66.4	151	US-09-741-873C-4	Sequence 4, Appl
2	439	56.4	131	US-09-741-873C-2	Sequence 2, Appl
3	92	11.8	1327	PCT-US04-02338-49	Sequence 7608, Ap
4	91.5	11.7	443	US-10-100-683-7608	Sequence 45603, A
5	90.5	11.6	234	US-10-767-701-45603	Sequence 63, Appl
6	85	10.9	956	US-10-093-037A-63	Sequence 10326, A
7	82	10.5	386	US-10-100-683-10326	Sequence 10327, A
8	82	10.5	386	US-10-100-683-10327	Sequence 26, Appl
9	81.5	10.5	1871	PCT-US03-02038-26	Sequence 314, App
10	78.5	10.1	430	US-10-451-467A-314	Sequence 1755, Ap
11	77	9.9	351	PCT-US04-05654-1155	Sequence 40, Appl
12	76	9.8	250	US-10-779-461-40	Sequence 590, App
13	76	9.8	321	PCT-US04-05654-590	Sequence 299, App
14	75	9.8	321	US-10-771-241-299	Sequence 2748, Ap
15	75	9.6	339	PCT-US04-05654-2748	Sequence 591, App
16	75	9.6	424	US-10-045-674A-591	Sequence 527, App
17	75	9.6	533	US-10-045-674A-527	Sequence 47075, A
18	74.5	9.6	1160	US-10-603-150-2	Sequence 47075, A
19	74	9.5	1201	US-10-767-701-47075	Sequence 6, Appl
20	74	9.5	250	US-10-779-461-6	Sequence 40104, A
21	73	9.4	205	US-10-767-701-40104	Sequence 41, Appl
22	73	9.4	246	US-10-779-461-41	Sequence 5818, Ap
23	72.5	9.3	318	US-10-417-884A-5818	Sequence 594, App
24	72	9.2	434	US-10-045-674A-594	Sequence 42696, A
25	71.5	9.2	179	US-10-767-701-42696	Sequence 32417, A
26	71.5	9.2	251	US-10-767-701-32417	

27	71.5	9.2	471	6	US-10-746-795A-22	Sequence 22, Appl
28	71	9.1	125	6	US-10-767-701-58970	Sequence 58970, A
29	71	9.1	310	6	US-10-786-850-4	Sequence 4, Appl
30	71	9.1	310	6	US-10-695-499-192	Sequence 192, App
31	71	9.1	319	6	US-10-695-499-68	Sequence 68, Appl
32	71	9.1	364	1	PCT-US04-05654-1756	Sequence 1756, Ap
33	71	9.1	619	6	US-10-695-499-70	Sequence 70, Appl
34	70.5	9.1	273	6	US-10-627-556-392	Sequence 392, App
35	70.5	9.1	273	6	US-10-627-556-396	Sequence 396, App
36	70.5	9.1	353	6	US-10-627-556-400	Sequence 45455, A
37	70.5	9.1	505	6	US-10-627-556-400	Sequence 400, App
38	70.5	9.1	505	6	US-10-627-556-402	Sequence 402, App
39	70.5	9.1	521	6	US-10-627-556-398	Sequence 398, App
40	70.5	9.1	1448	1	PCT-US04-05654-2086	Sequence 2086, Ap
41	70	9.0	175	6	US-10-767-701-35046	Sequence 35046, A
42	70	9.0	179	6	US-10-767-701-35342	Sequence 35342, A
43	70	9.0	199	6	US-10-767-701-51583	Sequence 51583, A
44	70	9.0	245	6	US-10-779-461-59	Sequence 59, Appl
45	70	9.0	574	1	PCT-US04-05654-334	Sequence 334, App

ALIGNMENTS

```
RESULT 1
US-09-741-873C-4
; Sequence 4, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Olsehn, Staffan
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873C-4
```

Query Match 66.4%; Score 517; DB 5; Length 151;

Best local similarity 66.2%; Pred. No. 1.2e+38;

Matches 100; Conservative 22; Mismatches 29; Indels 0; Gaps 0;

```
QY 1 MKLLKVAAPAAIVSGSALAGVPPQGGGNGHNGSSGPDSTLSIYQGSANALALQ 60
   |||||
DB 1 MKLLKVAAPAAIVSGSALAGVPPQGGGNGHNGSSGPDSTLSIYQGSANALALQ 60
   |||||
QY 61 SPARKSETTITGCGNGADVGAGNSITETLTONGFRNATTDONAKNSDITVOYGG 120
   |||||
DB 61 TPAKNSDLTITGCGNGADVGAGNSITETLTONGFRNATTDONAKNSDITVOYGG 120
   |||||
QY 121 NNAALVNDQVTRVTVTHMAHANNATANY 151
   |||||
DB 121 GNGAAVDQVTRVTVTHMAHANNATANY 151
   |||||
```

```

RESULT 2
US-09-741-873C-2
; Sequence 2, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Stafan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8601723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 2
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873C-2

Query Match
Best Local Similarity 56.4%; Score 439; DB 5; Length 131;
Matches 82; Conservative 21; Mismatches 28; Indels 0; Gaps 0;

QY 21 GVPPWGGGNGGNGSSGPDSTISTYQGSANALALQSDARKSETTTQSGYGAD 80
DB 1 GVPPGGGNGGNGGNNSSGPNSEINTYQGGGNSALALQTDARNSDITITQHGCGAGAD 60
QY 81 VGGAGNSTIELTONGFRNNATIDQNNAKNSDITVGVYGGNNAALVNDQLVTEVTHEM 140
DB 61 VGGGSDSSIDITQGFNSATLDQNNKNSSEMYKQFGGNGAADVDTASNSVNTQV 120
QY 141 AHAANATLANQY 151
DB 121 GFNNATAHOY 131

RESULT 3
PCT-US04-02338-49
; Sequence 49, Application PC/TUS0402338
; GENERAL INFORMATION:
; APPLICANT: EXELIXIS, INC.
; TITLE OF INVENTION: MAPCKs AS MODIFIERS OF THE APC AND AXIN PATHWAYS AND METHODS OF
; FILE REFERENCE: EX04-003C-PC
; CURRENT APPLICATION NUMBER: PCT/US04/02338
; PRIOR FILING DATE: 2004-01-28
; PRIOR APPLICATION NUMBER: US60/443,484
; PRIOR FILING DATE: 2003-01-29
; PRIOR APPLICATION NUMBER: US60/447,358
; PRIOR FILING DATE: 2003-02-11
; PRIOR APPLICATION NUMBER: US60/461,789
; PRIOR FILING DATE: 2003-04-10
; PRIOR APPLICATION NUMBER: US60/470,684
; PRIOR FILING DATE: 2003-05-14
; PRIOR APPLICATION NUMBER: US60/479,650
; PRIOR FILING DATE: 2003-06-19
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 49
; LENGTH: 1327

```

```

; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US04-02338-49

Query Match
Best Local Similarity 11.8%; Score 92; DB 1; Length 1327;
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

QY 6 VAAFAI-VVSGSALAGVPQWGGGNGGNGSSGPDSTISTYQGSANALALQSDAR 64
DB 99 VAAFPVPAVSTSSAGVAPNPAGSGSNNSSPTSS-SSSSPSSPGSSLAEPSEA 157
QY 65 KSEITIT---QSGYGADVGGAGNSTIELTONG--FRNNATIQMNAKSDI 113
DB 158 GVSSTAPLPGGAGPGTGVPVAVSALRELEACRNGDVSRKRLVDAANVAKDM 212

RESULT 4
US-10-100-683-7608
; Sequence 7608, Application US/10100683
; GENERAL INFORMATION:
; APPLICANT: Rosen, et al.
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: P8900
; CURRENT APPLICATION NUMBER: US/10/100,683
; PRIOR FILING DATE: 2002-03-19
; PRIOR APPLICATION NUMBER: US 60/040,162
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: US 60/043,576
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: US 60/047,601
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: US 60/056,845
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/043,580
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: US 60/047,599
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: US 60/056,664
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/043,314
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: US 60/047,632
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: US 60/056,892
; PRIOR FILING DATE: 1997-08-22
; Remaining Prior Application data removed - See file wrapper or PALM.
; NUMBER OF SEQ ID NOS: 13468
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 7608
; LENGTH: 443
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-100-683-7608

Query Match
Best Local Similarity 11.7%; Score 91.5; DB 6; Length 443;
Matches 33; Conservative 18; Mismatches 43; Indels 33; Gaps 6;

QY 21 GVPPQ---WGGGNGGNGSSGPDSTISTYV--YGSANALALQSDARKSETTTQSGY 75
DB 185 GNPQGAPWGGG--NGGPPNFGTITQGAIVAPGYSVRA-----SNQNEGTNPBPSSG 237
QY 76 GNGAD---VGGAGNSTIELTONGFRNNATI-----DQNNAKNSDIT 114
DB 238 GGGSSNSGGSGSSGSGSGSNGNNNGSSGSGSGSGSRDGSSESSNGSSTGSS 297
QY 115 VGQYGN 121
DB 298 SGNHGS 304

RESULT 5

```



```

; CURRENT FILING DATE: 2002-03-19
; PRIOR APPLICATION NUMBER: US 60/040,162
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: US 60/043,576
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: US 60/047,601
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: US 60/056,845
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/043,580
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: US 60/047,599
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: US 60/056,664
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/043,314
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: US 60/047,632
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: US 60/056,892
; PRIOR FILING DATE: 1997-08-22
; Remaining Prior Application data removed - See file wrapper or PALM.
; NUMBER OF SEQ ID NOS: 13468
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10327
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-100-683-10327

```

```

Query Match      10.5%; Score 82; DB 6; Length 386;
Best Local Similarity 29.1%; Pred. No. 4.1;
Matches 32; Conservative 18; Mismatches 44; Indels 16; Gaps 6;

```

```

Qy 21 GVPQ---WGCGNNNGSGSPDSTLSIYQ--YGSNAAALALQSDARKSETTITGQGY 75
Db 185 GNPFCAPWGGGQ--NGGPPNFGTNTQGAVPYGVRA-----SNQNECTPMPPEGS 237
Qy 76 GNGAD---VGCGADNSTIELTQNGRNATIDQWNAKNSDITVGOYGANN 122
Db 228 GGGSSNSGGGSGSGSSGSSGSGSNGNNGSSSGSSSS--SGSSSGSS 286

```

```

RESULT 9
PCT-US03-02038-26
; Sequence 26, Application PC/TUS0302038
; GENERAL INFORMATION:
; APPLICANT: Mayo Foundation for Medical Education and Research
; TITLE OF INVENTION: Polycystic Kidney Disease Nucleic Acids
; FILE REFERENCE: 07039/386W01
; CURRENT APPLICATION NUMBER: PCT/US03/02038
; CURRENT FILING DATE: 2003-01-23
; PRIOR APPLICATION NUMBER: 60/351,110
; PRIOR FILING DATE: 2002-01-23
; NUMBER OF SEQ ID NOS: 221
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 1871
; TYPE: PRT
; ORGANISM: Mus musculus
; PCT-US03-02038-26

```

```

Query Match      10.5%; Score 81.5; DB 1; Length 1871;
Best Local Similarity 23.0%; Pred. No. 31;
Matches 28; Conservative 13; Mismatches 42; Indels 39; Gaps 3;

```

```

Qy 26 WGGGNNHNGGSSGPDSTLSIYQGSNAAALALQSDARKSETTITGSGYNGADVGQGA 85
Db 17 WCAADSHKGSSETI-PKYTEVIEPKYGSINGATRL-----TIKGEFSQASOFNYGA 66
Qy 86 DNSTIELTQNGRNATIDQWNAKNSDITVGOYGNNAAALVNDQTLRVVTHEMAHANN 145

```

```

Db 67 DNTEL-----GNHVQVVSFQSIITCDVEKDSHSTQ 97
Qy 146 AT 147
Db 98 IT 99

```

```

RESULT 10
US-10-451-467A-314
; Sequence 314, Application US/10451467A
; GENERAL INFORMATION:
; APPLICANT: CONTRERAS, ROLAND HENRI
; APPLICANT: EBERHARDT, INES
; APPLICANT: LUTTEN, WALTER HERMAN MARIA LOUIS
; APPLICANT: REEKMAN, RIEKA JOSEPHINA
; TITLE OF INVENTION: BAX-RESPONSIVE GENES FOR DRUG TARGET IDENTIFICATION IN
; FILE REFERENCE: JAB-1667
; CURRENT APPLICATION NUMBER: US/10/451,467A
; CURRENT FILING DATE: 2003-06-19
; PRIOR APPLICATION NUMBER: EP 00870318.3
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: EP 01870002.1
; PRIOR FILING DATE: 2001-01-04
; PRIOR APPLICATION NUMBER: EP 01870003.9
; PRIOR FILING DATE: 2001-01-09
; NUMBER OF SEQ ID NOS: 732
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 314
; LENGTH: 430
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
; US-10-451-467A-314

```

```

Query Match      10.1%; Score 78.5; DB 6; Length 430;
Best Local Similarity 28.4%; Pred. No. 9.5;
Matches 27; Conservative 12; Mismatches 39; Indels 17; Gaps 5;

```

```

Qy 29 GGNHNGGSSGPDSTLSIYQGS--ANAAALALQSDARKSETTITGSGYNGADVGQGAN 87
Db 89 GSNNDNSYGSNNDS-----YGSNNDSYGSNNDSYGSNNKKSXYSYGSNNDSYGSNN 142

```

```

Qy 88 STIELTQNGRNATIDQWNAKNSDITVGOYGANN 122
Db 143 NN---DSYGSNN---DSYGSNNND---SYGSNN 167

```

```

RESULT 11
PCT-US04-05654-1755
; Sequence 1755, Application PC/TUS0405654
; GENERAL INFORMATION:
; APPLICANT: Sherman, Bradley K
; APPLICANT: Rieckmann, Jose Luis
; APPLICANT: Liang, Cal-Zhong
; APPLICANT: Heard, Jacqueline E
; APPLICANT: Haake, Volker
; APPLICANT: Creelman, Robert A
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Adam, Luc J
; APPLICANT: Reuder, T. Lynne
; APPLICANT: Keddie, James
; APPLICANT: Dubell III, Arnold N
; APPLICANT: Pinada, Omaira
; APPLICANT: Repetti, Peter
; APPLICANT: Guttererson, Neal
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Brown, Pierre E
; APPLICANT: Kumamoto, Roderick W
; APPLICANT: Pilgrim, Marsha L
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES IN PLANTS
; FILE REFERENCE: MBI-0047 PCT
; CURRENT APPLICATION NUMBER: PCT/US04/05654

```

```

; CURRENT FILING DATE: 2004-03-03
; PRIOR APPLICATION NUMBER: 10/374,780
; PRIOR FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: 10/675,852
; PRIOR FILING DATE: 2003-09-30
; NUMBER OF SEQ ID NOS: 2950
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1755
; LENGTH: 351
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Orthologous to GI1794
PCT-US04-05654-1755

Query Match          9.8%; Score 77; DB 1; Length 351;
Best Local Similarity 31.2%; Pred. No. 10;
Matches 35; Conservative 15; Mismatches 42; Indels 20; Gaps 6;

QY 17 SALAGVVPQWGGGNNHGGG---NSSGPDSTLSIYOGSANAALALQSDARKSETTTTQ 72
DB 24 SALARVVAAGGGGGGDDGDGMAMWSFSPSS-----AAAAAARGVRRRRRREGQAMHEL 77
QY 73 SCYG-NCADYGQAGDNSTIELTONGFRNNATTIDQMAKNSDITVGYGGNNA 123
DB 78 AGYACGAPSPREFAGSBOSSDTOSA--SAATMDEHHS-----PVG--GGGNA 120

RESULT 12
US-10-779-461-40
; Sequence 40, Application US/10779461
; GENERAL INFORMATION:
; APPLICANT: Morton, Philip A
; TITLE OF INVENTION: ANTIBODIES TO C-MET FOR THE TREATMENT OF CANCERS
; FILE REFERENCE: 00980/1
; CURRENT APPLICATION NUMBER: US/10/779,461
; CURRENT FILING DATE: 2004-02-13
; PRIOR APPLICATION NUMBER: 60/447,073
; PRIOR FILING DATE: 2003-02-13
; NUMBER OF SEQ ID NOS: 161
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 40
; LENGTH: 250
; TYPE: PRT
; ORGANISM: artificial
; FEATURE:
; OTHER INFORMATION: phage display generated human antibody
US-10-779-461-40

Query Match          9.8%; Score 76; DB 6; Length 250;
Best Local Similarity 30.6%; Pred. No. 8.1;
Matches 26; Conservative 8; Mismatches 31; Indels 20; Gaps 3;

QY 9 FAALVVGSLAGVPMWG-----GGNNHGGNSSGPDSTLSIYOGSANAALAL 59
DB 95 YCAISPLRGITADVDFVDGGLVTVSSGGGSGGSGGSGG-----GSAQSALIQ 144
QY 60 QSDARKSE-TTITQSGYGNADVGQ 83
DB 145 PASASGSPGOSTITICTGTSSDIGR 169

RESULT 13
PCT-US04-05654-590
; Sequence 590, Application PC/TUS0405654
; GENERAL INFORMATION:
; APPLICANT: Sherman, Bradley K
; APPLICANT: Riechmann, Jose Luis
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Heard, Jacqueline E
; APPLICANT: Haake, Volker
; APPLICANT: Creelman, Robert A
; APPLICANT: Ratcliffe, Oliver

```

```

; APPLICANT: Adam, Luc J
; APPLICANT: Reuber, T. Lynne
; APPLICANT: Keddie, James
; APPLICANT: Dubell III, Arnold N
; APPLICANT: Pineda, Omaira
; APPLICANT: Repetti, Peter
; APPLICANT: Century, Karen
; APPLICANT: Guttersen, Neal
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Brown, Pierre E
; APPLICANT: Kumamoto, Roderick W
; APPLICANT: Pilgrim, Marsha L
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES IN PLANTS
; FILE REFERENCE: MBI-0047 PCT
; CURRENT APPLICATION NUMBER: PCT/US04/05654
; CURRENT FILING DATE: 2004-03-03
; PRIOR APPLICATION NUMBER: 10/374,780
; PRIOR FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: 10/675,852
; PRIOR FILING DATE: 2003-09-30
; NUMBER OF SEQ ID NOS: 2950
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 590
; LENGTH: 321
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: G3384 Orthologous to G256
PCT-US04-05654-590

Query Match          9.8%; Score 76; DB 1; Length 321;
Best Local Similarity 24.4%; Pred. No. 11;
Matches 29; Conservative 18; Mismatches 42; Indels 30; Gaps 5;

QY 10 AAIVVGSALAGVPMW---GGGNNHGGNSSGPDSTLSIYOGSANAALALQSDARKS 66
DB 199 AAYASSADNIARLLQGMKPPGG---GGNGGPPAS-----GSTTTATTQOQPOQS 248
QY 67 ETTITQSGYGNADVGQAGDNSTIELTONGFRNNATTID--QWNAKNSDITVGYGGNNA 123
DB 249 -----GEGALASASASOSGAAAAAATNOTPECTETSKATGGAGGPA 292

RESULT 14
US-10-771-241-299
; Sequence 299, Application US/10771241
; GENERAL INFORMATION:
; APPLICANT: Zykkind, Judith
; APPLICANT: Forsyth, R. Allyn
; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
; FILE REFERENCE: ELITRA 001C1
; CURRENT APPLICATION NUMBER: US/10/771,241
; CURRENT FILING DATE: 2004-02-03
; PRIOR APPLICATION NUMBER: 09/492,709
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: 60/117,405
; PRIOR FILING DATE: 1999-01-27
; NUMBER OF SEQ ID NOS: 485
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 299
; LENGTH: 382
; TYPE: PRT
; ORGANISM: E. Coli
US-10-771-241-299

Query Match          9.8%; Score 76; DB 6; Length 382;
Best Local Similarity 27.7%; Pred. No. 14;
Matches 38; Conservative 13; Mismatches 50; Indels 36; Gaps 9;

QY 30 GNNHGGNSSGPDSTLSI-YQ--YGSANAALALQSDARKSETTTTQSG-----YGNADVG 82
DB 158 GNTHGNTSSGTSGYSLNRYAGYNTVGYRSRGS--SQIYGMGGIHAADGITRG 215

```

QY 83 0-----GADNSTELTONGFRNNATIDOMNANSDITVQYGGANNAAL----- 125
Db 216 QPGLDGMVLYKAPGADNVKIE-NQTHITD-----WRGYAILPFATEYRENNVALNANSL 269
QY 126 ---VNYDQLVTRVY-TH 138
Db 270 ADNVELEDETIVTVIPTH 286

RESULT 15

PCT-US04-05654-2748
; Sequence 2748, Application PC/TUS0405654
; GENERAL INFORMATION:
; APPLICANT: Sherman, Bradley K
; APPLICANT: Riechmann, Jose Luis
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Heard, Jacqueline E
; APPLICANT: Haake, Volker
; APPLICANT: Creelman, Robert A
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Adam, Luc J
; APPLICANT: Reuber, T. Lynne
; APPLICANT: Keddie, James
; APPLICANT: Dubell III, Arnold N
; APPLICANT: Pineda, Omaira
; APPLICANT: Repetti, Peter
; APPLICANT: Century, Karen
; APPLICANT: Guterson, Neal
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Brown, Pierre E
; APPLICANT: Kumamoto, Roderick W
; APPLICANT: Pilgrim, Marsha L
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES IN PLANTS
; FILE REFERENCE: MRI-0047 PCT
; CURRENT APPLICATION NUMBER: PCT/US04/05654
; CURRENT FILING DATE: 2004-03-03
; PRIOR APPLICATION NUMBER: 10/374,780
; PRIOR FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: 10/675,852
; PRIOR FILING DATE: 2003-09-30
; NUMBER OF SEQ ID NOS: 2950
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2748
; LENGTH: 399
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: G1882
PCT-US04-05654-2748

Query Match 9.6%; Score 75; DB 1; Length 399;
Best Local Similarity 26.7%; Pred. No. 18;
Matches 32; Conservative 14; Mismatches 58; Indels 16; Gaps 4;
QY 27 GCGGNHNGGNSGSGPDSLSTIYQ---GSANAALALQSDARKSETTTGSGYG----- 76
Db 155 GGGGGSTSSGNSKSDSATSNQYHHRAMANNQMGPPSSSSLSLSLVNAGLIPGHDH 214
QY 77 ---NCADYVGAGADNSTIELTQN-GFRNNATIDQW--NAKNSDITVQYGGANNAALVNYDQ 130
Db 215 NSNNNNIIGLSSLPPLKMPPLDFTDFTLQYGAVASAPSYHIGGSSGGAAALLNGFDQ 274

Search completed: March 11, 2004, 19:14:15
Job time : 6.3 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:33:51 ; Search time 171.3 Seconds
(without alignments)
860.386 Million cell updates/sec

Title: US-09-543-407-14

Perfect score: 1 MCLKVAFAAIVVSGSALA.....HASVWVRQVGFNNATANQY 151

Sequence: 1 MCLKVAFAAIVVSGSALA.....HASVWVRQVGFNNATANQY 151

Scoring table: BLOSUM62
Gap 10.0 , Gapext 0.5

Searched: 6019581 seqs, 976053577 residues

Total number of hits satisfying chosen parameters: 6019581

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Parents AA Main:*

1:	/cgn2_6/ptodata/2/paa/PTCUS_COMB.pep.*
2:	/cgn2_6/ptodata/2/paa/US06_COMB.pep.*
3:	/cgn2_6/ptodata/2/paa/US07_COMB.pep.*
4:	/cgn2_6/ptodata/2/paa/US08_COMB.pep.*
5:	/cgn2_6/ptodata/2/paa/US082_COMB.pep.*
6:	/cgn2_6/ptodata/2/paa/US083_COMB.pep.*
7:	/cgn2_6/ptodata/2/paa/US085_COMB.pep.*
8:	/cgn2_6/ptodata/2/paa/US085_COMB.pep.*
9:	/cgn2_6/ptodata/2/paa/US085_COMB.pep.*
10:	/cgn2_6/ptodata/2/paa/US085_COMB.pep.*
11:	/cgn2_6/ptodata/2/paa/US087_COMB.pep.*
12:	/cgn2_6/ptodata/2/paa/US088_COMB.pep.*
13:	/cgn2_6/ptodata/2/paa/US089_COMB.pep.*
14:	/cgn2_6/ptodata/2/paa/US090_COMB.pep.*
15:	/cgn2_6/ptodata/2/paa/US091_COMB.pep.*
16:	/cgn2_6/ptodata/2/paa/US092_COMB.pep.*
17:	/cgn2_6/ptodata/2/paa/US093_COMB.pep.*
18:	/cgn2_6/ptodata/2/paa/US094_COMB.pep.*
19:	/cgn2_6/ptodata/2/paa/US095_COMB.pep.*
20:	/cgn2_6/ptodata/2/paa/US096_COMB.pep.*
21:	/cgn2_6/ptodata/2/paa/US097A_COMB.pep.*
22:	/cgn2_6/ptodata/2/paa/US097B_COMB.pep.*
23:	/cgn2_6/ptodata/2/paa/US098_COMB.pep.*
24:	/cgn2_6/ptodata/2/paa/US099A_COMB.pep.*
25:	/cgn2_6/ptodata/2/paa/US099B_COMB.pep.*
26:	/cgn2_6/ptodata/2/paa/US099B_COMB.pep.*
27:	/cgn2_6/ptodata/2/paa/US101_COMB.pep.*
28:	/cgn2_6/ptodata/2/paa/US101_COMB.pep.*
29:	/cgn2_6/ptodata/2/paa/US103_COMB.pep.*
30:	/cgn2_6/ptodata/2/paa/US104_COMB.pep.*
31:	/cgn2_6/ptodata/2/paa/US106_COMB.pep.*
32:	/cgn2_6/ptodata/2/paa/US107_COMB.pep.*
33:	/cgn2_6/ptodata/2/paa/US60_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length DB	ID	Description
------------	-------------	-----------	----	-------------

1	775	100.0	151	19	US-09-543-407-14	Sequence 14, Appl
2	714	92.1	151	19	US-09-543-407-24	Sequence 24, Appl
3	696	89.8	151	19	US-09-543-407-12	Sequence 12, Appl
4	693	89.4	151	19	US-09-543-407-5	Sequence 5, Appl
5	692	89.3	151	6	US-08-233-6424-57	Sequence 57, Appl
6	659	85.0	151	6	US-09-543-407-30	Sequence 30, Appl
7	612	79.0	151	19	US-09-543-407-26	Sequence 26, Appl
8	611	78.8	151	19	US-09-543-407-18	Sequence 18, Appl
9	609	78.6	151	19	US-09-543-407-20	Sequence 20, Appl
10	606	78.2	131	19	US-09-543-407-20	Sequence 20, Appl
11	603	77.8	151	19	US-09-543-407-28	Sequence 28, Appl
12	602	77.7	151	19	US-09-543-407-22	Sequence 22, Appl
13	578	74.6	151	19	US-09-543-407-16	Sequence 16, Appl
14	528	68.1	151	13	US-09-543-407-7	Sequence 7, Appl
15	525	67.7	151	13	US-08-978-878-4	Sequence 4, Appl
16	525	67.7	151	21	US-09-741-873B-4	Sequence 4, Appl
17	523	67.5	151	33	US-60-352-946-2	Sequence 2, Appl
18	523	67.5	151	33	US-60-444-371-2	Sequence 2, Appl
19	514	66.3	120	6	US-08-233-6424-55	Sequence 55, Appl
20	474	61.2	109	19	US-09-543-407-34	Sequence 34, Appl
21	472	60.9	158	16	US-09-252-691C-5834	Sequence 5834, Ap
22	472	60.9	158	16	US-09-252-691C-5834	Sequence 5834, Ap
23	472	60.9	158	30	US-10-417-886-5834	Sequence 5834, Ap
24	447	57.7	131	13	US-08-978-878-2	Sequence 2, Appl
25	447	57.7	131	21	US-09-543-407-37	Sequence 37, Appl
26	347	44.8	68	19	US-09-543-407-35	Sequence 35, Appl
27	343	44.3	109	19	US-09-543-407-32	Sequence 32, Appl
28	243.5	31.4	70	19	US-09-543-407-32	Sequence 32, Appl
29	237	30.6	48	19	US-09-543-407-39	Sequence 39, Appl
30	115.5	14.9	186	16	US-09-252-691C-5833	Sequence 5833, Ap
31	115.5	14.9	186	16	US-09-252-691C-5833	Sequence 5833, Ap
32	115.5	14.9	186	10	US-10-417-886-5833	Sequence 8, Appl
33	104.5	13.5	151	19	US-09-543-407-8	Sequence 8, Appl
34	104	13.4	445	29	US-10-369-493-20638	Sequence 20638, A
35	104	13.4	445	33	US-60-360-039-20638	Sequence 20638, A
36	98.5	12.7	151	19	US-09-543-407-6	Sequence 6, Appl
37	98	12.6	210	28	US-10-219-999-43013	Sequence 43013, A
38	98	12.6	210	30	US-10-425-114-53421	Sequence 53421, A
39	98	12.6	210	30	US-10-425-114-53421	Sequence 53421, A
40	98	12.6	210	33	US-60-312-544-6755	Sequence 6755, Ap
41	98	12.6	210	33	US-60-324-109-30520	Sequence 30520, A
42	98	12.6	211	30	US-10-424-599-245046	Sequence 245046, A
43	95	12.3	186	28	US-10-282-122A-49412	Sequence 49412, A
44	92	11.9	673	20	US-09-678-411-8	Sequence 8, Appl
45	92	11.9	949	20	US-09-678-411-10	Sequence 10, Appl

ALIGNMENTS

RESULT 1
US-09-543-407-14
Sequence 14, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543,407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ. ID NOS.: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 14
LENGTH: 151
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.

US-09-543-407-14

Query Match 100.0%; Score 775; DB 19; Length 151;
Best Local Similarity 100.0%; Pred. No. 6, 1e-75;
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLKVAFAFAIVVSGSALAGVVPQWGGGNNHNGSGSPDSTLSIYQGSANALALQ 60
DB 1 MKLLKVAFAFAIVVSGSALAGVVPQWGGGNNHNGSGSPDSTLSIYQGSANALALQ 60
QY 61 SDARKSETTITGSGYGNADVGGADNSTIELTONGFRNNATTIDQNAKNSDITVGYQYDQ 120
DB 61 SDARKSETTITGSGYGNADVGGADNSTIELTONGFRNNATTIDQNAKNSDITVGYQYDQ 120
QY 121 LVTRVVTHEMAHASVWVRQVGFNNATANQY 151
DB 121 LVTRVVTHEMAHASVWVRQVGFNNATANQY 151

RESULT 2

US-09-543-407-24

; Sequence 24, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543.407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 151
; TYPE: PRF
; ORGANISM: Artificial Sequence
; FEATURE: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-24

Query Match 92.1%; Score 714; DB 19; Length 151;
Best Local Similarity 91.1%; Pred. No. 2, 4e-68;
Matches 144; Conservative 0; Mismatches 0; Indels 14; Gaps 2;

QY 1 MKLLKVAFAFAIVVSGSALAGVVPQWGGGNNHNGSGSPDSTLSIYQGSANALALQ 60
DB 1 MKLLKVAFAFAIVVSGSALAGVVPQWGGGNNHNGSGSPDSTLSIYQGSANALALQ 60
QY 61 SDARKSETTITGSGYGNADVGGADNSTIELTONGFRNNATTIDQNAKNSDITVGYQYDQ 120
DB 61 SDARKSETTITGSGYGNADVGGADNSTIELTONGFRNNATTIDQNAKNSDITVGYQYDQ 120
QY 121 LVTRVVTHEMAHASVWVRQVGFNNATANQY 151
DB 121 LVTRVVTHEMAHASVWVRQVGFNNATANQY 151

RESULT 3

US-09-543-407-12

; Sequence 12, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543.407

CURRENT FILING DATE: 2000-04-05

; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 151
; TYPE: PRF
; ORGANISM: Artificial Sequence
; FEATURE: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-12

Query Match 89.8%; Score 696; DB 19; Length 151;
Best Local Similarity 87.6%; Pred. No. 2, 1e-66;
Matches 141; Conservative 0; Mismatches 0; Indels 20; Gaps 2;

QY 1 MKLLKVAFAFAIVVSGSALAGVVPQWGGGNNHNGSGSPDSTLSIYQGSANALALQ 60
DB 1 MKLLKVAFAFAIVVSGSALAGVVPQWGGGNNHNGSGSPDSTLSIYQGSANALALQ 60
QY 61 SDARKSETTITGSGYGNADVGGADNSTIELTONGFRNNATTIDQNAKNSDITVGYQYDQ 117
DB 61 SDARKSETTITGSGYGNADVGGADNSTIELTONGFRNNATTIDQNAKNSDITVGYQYDQ 120
QY 118 -----YDQLVTRVVTHEMAHASVWVRQVGFNNATANQY 151
DB 121 NNAALVNTQVTLVTRVVTHEMAHASVWVRQVGFNNATANQY 151

RESULT 4

US-09-543-407-5

; Sequence 5, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543.407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 151
; TYPE: PRF
; ORGANISM: Salmonella enteritidis
US-09-543-407-5

Query Match 89.4%; Score 693; DB 19; Length 151;
Best Local Similarity 90.7%; Pred. No. 4, 4e-66;
Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 1 MKLLKVAFAFAIVVSGSALAGVVPQWGGGNNHNGSGSPDSTLSIYQGSANALALQ 60
DB 1 MKLLKVAFAFAIVVSGSALAGVVPQWGGGNNHNGSGSPDSTLSIYQGSANALALQ 60
QY 61 SDARKSETTITGSGYGNADVGGADNSTIELTONGFRNNATTIDQNAKNSDITVGYQYDQ 120
DB 61 SDARKSETTITGSGYGNADVGGADNSTIELTONGFRNNATTIDQNAKNSDITVGYQYDQ 120
QY 121 LVTRVVTHEMAHASVWVRQVGFNNATANQY 151
DB 121 NNAALVNTQVTLVTRVVTHEMAHASVWVRQVGFNNATANQY 151

RESULT 5

US-08-233-642A-57

; Sequence 57, Application US/08233642A
; GENERAL INFORMATION:
; APPLICANT: Kay, William W.

```

; APPLICANT: Collinson, S. Karen
; APPLICANT: Clouthier, Sharon C.
; APPLICANT: Doran, James L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
; TITLE OF INVENTION: BASED VACCINES
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,642A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE//DOCKET NUMBER: 920043.403C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4800
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDNBERRY
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 151 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-233-642A-57

Query Match          89.3%; Score 692; DB 6; Length 151;
Best Local Similarity 90.7%; Pred. No. 5.6e-66;
Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 1 MKLLKVAFAFAIIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANAALALQ 60
DB 1 MKLLKVAFAFAIIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANAALALQ 60
QY 61 SDAKSETTTTQSGYGADVGAGADNSTLTONGFRNNATTIDQNAKXSDITVGOYDQ 120
DB 61 SDAKSETTTTQSGYGADVGAGADNSTLTONGFRNNATTIDQNAKXSDITVGOYDQ 120
QY 121 LVTRVVTHEMAHSAVMVRQVFGNNATANY 151
DB 121 NNPAALVNGTASDSSVMVRQVFGNNATANY 151

RESULT 6
US-09-543-407-30
; Sequence 30, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 151

```

```

; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding P73 from Gpe3 of Leishmania major.
; US-09-543-407-30

Query Match          85.0%; Score 659; DB 19; Length 151;
Best Local Similarity 81.3%; Pred. No. 2.1e-62;
Matches 136; Conservative 0; Mismatches 0; Indels 30; Gaps 2;

QY 1 MKLLKVAFAFAIIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANAALALQ 60
DB 1 MKLLKVAFAFAIIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANAALALQ 60
QY 61 SDAKSETTTTQSGYGADVGAGADNSTLTONGFRNNATTIDQNAKXSDITVGOYDQ 120
DB 61 SDAKSETTTTQSGYGADVGAGADNSTLTONGFRNNATTIDQNAKXSDITVGOYDQ 120
QY 121 LVTRVVTHEMAHSAVMVRQVFGNNATANY 151
DB 121 NNPAALVNGTASDSSVMVRQVFGNNATANY 151

RESULT 7
US-09-543-407-26
; Sequence 26, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding P73 from Gpe3 of Leishmania major.
; US-09-543-407-26

Query Match          79.0%; Score 612; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 2.5e-57;
Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAFAFAIIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANAALALQ 60
DB 1 MKLLKVAFAFAIIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANAALALQ 60
QY 61 SDAKSETTTTQSGYGADVGAGADNSTLTONGFRNNATTIDQNAKXSDITVGOYDQ 120
DB 61 SDAKSETTTTQSGYGADVGAGADNSTLTONGFRNNATTIDQNAKXSDITVGOYDQ 120
QY 121 LVTRVVTHEMAHSAVMVRQVFGNNATANY 151
DB 121 NNPAALVNGTASDSSVMVRQVFGNNATANY 151

RESULT 8
US-09-543-407-18
; Sequence 18, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.

```

```

; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-18

```

```

Query Match      78.8%; Score 611; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 3.2e-57;
Matches 123; Conservative 6; Mismatches 22; Indels 0; Gaps 0;

```

```

Qy 1 MKLKVAFPAIVSGSALAGVPPWGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
Db 1 MKLKVAFPAIVSGSALAGVPPWGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
Qy 61 SDARKSETTITGSGYNGADVGGADNSTLTONGFRNNATTIDQNNAKNSDITVGYDQ 120
Db 61 SDARKSETTITGSGYNGADVGGADNSTLTONGFRNNATTIDQNNAKNSDITVGYDQ 120
Qy 121 LVTRVVTHEMAHSAVMRVQVFGNNATANY 151
Db 121 NNALVNQTASDSVMRVQVFGNNATANY 151

```

```

RESULT 9
US-09-543-407-20
; Sequence 20, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-20

```

```

Query Match      78.6%; Score 609; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 5.2e-57;
Matches 123; Conservative 6; Mismatches 22; Indels 0; Gaps 0;

```

```

Qy 1 MKLKVAFPAIVSGSALAGVPPWGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
Db 1 MKLKVAFPAIVSGSALAGVPPWGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
Qy 61 SDARKSETTITGSGYNGADVGGADNSTLTONGFRNNATTIDQNNAKNSDITVGYDQ 120
Db 61 SDARKSETTITGSGYNGADVGGADNSTLTONGFRNNATTIDQNNAKNSDITVGYDQ 120

```

```

Qy 121 LVTRVVTHEMAHSAVMRVQVFGNNATANY 151
Db 121 NNALVNQTASDSVMRVQVFGNNATANY 151

```

```

RESULT 10
US-09-543-407-31
; Sequence 31, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Salmonella enteritidis
US-09-543-407-31

```

```

Query Match      78.2%; Score 606; DB 19; Length 131;
Best Local Similarity 89.3%; Pred. No. 9e-57;
Matches 117; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

```

```

Qy 21 GVPWQGGGNNHNGGNSGPDSTLSIYQGSANALALQSDARKSETTITGSGYNGAD 80
Db 1 GVPWQGGGNNHNGGNSGPDSTLSIYQGSANALALQSDARKSETTITGSGYNGAD 80
Qy 81 VGGADNSTLTONGFRNNATTIDQNNAKNSDITVGYDQLVTRVVTHEMAHSAVMRVQ 140
Db 61 VGGADNSTLTONGFRNNATTIDQNNAKNSDITVGYDQNNALVNQTASDSVMRVQ 120
Qy 141 GFGNNATANY 151
Db 121 GFGNNATANY 131

```

```

RESULT 11
US-09-543-407-28
; Sequence 28, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-28

```

```

Query Match      77.8%; Score 603; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 2.3e-56;
Matches 122; Conservative 4; Mismatches 25; Indels 0; Gaps 0;

```

```

Qy 1 MKLKVAFPAIVSGSALAGVPPWGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60

```


DB 1 MLKLVAAFAAIVSGSALAGVPPQGGGNNHGGSSGPDSTLSIYOGSANAALAQ 60
QY 61 SPARKSETTITGSGYNGADVGGADNSTIELTONGFRNNATIDOMNAKSDITVGOYDQ 120
DB 61 SPARKSETTITGSGYNGADVGGADNSTIELTONGFRNNATIDOMNAKSDITVGOYDQ 120
QY 121 LVTRVTHEMAHASVVRQVGFNNATANQY 151
DB 121 NNAALVNOTASDSVVRQVGFNNATANQY 151

RESULT 12
US-09-543-407-22
; Sequence 22, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 22
; LENGTH: 151
; TYPE: PRF
; ORGANISM: Artificial Sequence
; OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
US-09-543-407-22

Query Match 77.7%; Score 602; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 3e-56;
Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;

QY 1 MLKLVAAFAAIVSGSALAGVPPQGGGNNHGGSSGPDSTLSIYOGSANAALAQ 60
DB 1 MLKLVAAFAAIVSGSALAGVPPQGGGNNHGGSSGPDSTLSIYOGSANAALAQ 60
QY 61 SPARKSETTITGSGYNGADVGGADNSTIELTONGFRNNATIDOMNAKSDITVGOYDQ 120
DB 61 SPARKSETTITGSGYNGADVGGADNSTIELTONGFRNNATIDOMNAKSDITVGOYDQ 120
QY 121 LVTRVTHEMAHASVVRQVGFNNATANQY 151
DB 121 NNAALVNOTASDSVVRQVGFNNATANQY 151

RESULT 13
US-09-543-407-16
; Sequence 16, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 151
; TYPE: PRF
; ORGANISM: Artificial Sequence

FEATURE:
; OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of *Leishmania* major.
US-09-543-407-16

Query Match 74.6%; Score 578; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 1.2e-53;
Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;

QY 1 MLKLVAAFAAIVSGSALAGVPPQGGGNNHGGSSGPDSTLSIYOGSANAALAQ 60
DB 1 MLKLVAAFAAIVSGSALAGVPPQGGGNNHGGSSGPDSTLSIYOGSANAALAQ 60
QY 61 SPARKSETTITGSGYNGADVGGADNSTIELTONGFRNNATIDOMNAKSDITVGOYDQ 120
DB 61 SPARKSETTITGSGYNGADVGGADNSTIELTONGFRNNATIDOMNAKSDITVGOYDQ 120
QY 121 LVTRVTHEMAHASVVRQVGFNNATANQY 151
DB 121 NNAALVNOTASDSVVRQVGFNNATANQY 151

RESULT 14
US-09-543-407-7
; Sequence 7, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 151
; TYPE: PRF
; ORGANISM: *Escherichia coli*
US-09-543-407-7

Query Match 68.1%; Score 528; DB 19; Length 151;
Best Local Similarity 68.9%; Pred. No. 2.9e-48;
Matches 104; Conservative 20; Mismatches 27; Indels 0; Gaps 0;

QY 1 MLKLVAAFAAIVSGSALAGVPPQGGGNNHGGSSGPDSTLSIYOGSANAALAQ 60
DB 1 MLKLVAAFAAIVSGSALAGVPPQGGGNNHGGSSGPDSTLSIYOGSANAALAQ 60
QY 61 SPARKSETTITGSGYNGADVGGADNSTIELTONGFRNNATIDOMNAKSDITVGOYDQ 120
DB 61 TDARNSDLTITOHGGNGADVGGSDSDSIDLTORFGNSATLIDOMNGKNSMTYKQFGG 120
QY 121 LVTRVTHEMAHASVVRQVGFNNATANQY 151
DB 121 GNGAADVOTASNSVVRQVGFNNATANQY 151

RESULT 15
US-08-978-878-4
; Sequence 4, Application US/08978878
; GENERAL INFORMATION:
; APPLICANT: NORMARK, Scaffan
; APPLICANT: OLSEN, Arne
; TITLE OF INVENTION: FIBRONECTIN BINDING PROTEIN AS WELL AS ITS PREPARATION
; FILE REFERENCE: 012889-081
; CURRENT APPLICATION NUMBER: US/08/978,878
; CURRENT FILING DATE: 1997-11-26
; EARLIER APPLICATION NUMBER: SE 8801723-1
; EARLIER FILING DATE: 1988-05-06

EARLIER APPLICATION NUMBER: US 07/347,189
EARLIER FILING DATE: 1989-05-04
EARLIER APPLICATION NUMBER: US 07/789,437
EARLIER FILING DATE: 1991-11-06
EARLIER APPLICATION NUMBER: US 07/970,846
EARLIER FILING DATE: 1992-11-03
EARLIER APPLICATION NUMBER: US 08/187,865
EARLIER FILING DATE: 1994-01-28
EARLIER APPLICATION NUMBER: US 08/318,519
EARLIER FILING DATE: 1994-10-05
EARLIER APPLICATION NUMBER: US 08/495,959
EARLIER FILING DATE: 1995-06-28
NUMBER OF SEQ ID NOS: 10
SOFTWARE: Patencin Ver. 2.0
SEQ ID NO: 4
LENGTH: 151
TYPE: PRF
ORGANISM: Escherichia coli
US-08-978-878-4

Query Match 67.7%; Score 525; DB 13; Length 151;
Best Local Similarity 68.2%; Pred. No. 6.1e-48;
Matches 103; Conservative 21; Mismatches 27; Indels 0; Gaps 0;

QY 1 MKLKVAFPAIVTSGSALAGVPMWGGGNNHGGSSGPDSTLSTIYQGSANALALQ 60
DB 1 MKLKVAFPAIVTSGSALAGVPMWGGGNNHGGSSGPDSTLSTIYQGSANALALQ 60
QY 61 SPARKSETTITQSGYGADYVQGGADNSTIELTONGFERNATIIDQWNAKNSDITVGOYDQ 120
DB 61 TDARNSDLTIQHGGGADYVQGGSDSSIDLQRFGENSATLQDQWNGKNSMTVQFGG 120
QY 121 LVTRVYTHMAHMASVWQVGFNNATNAY 151
DB 121 GNGAAVDQTASNSVNTQVGFNNATAHY 151

Search completed: March 11, 2004, 19:13:10
Job time : 172.3 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:35:05 ; Search time 5.3 Seconds
(without alignments)
376.014 Million cell updates/sec

Title: US-09-543-407-14

Perfect score: 775

Sequence: 1 MLKLKVAFAIVGSGSALA.....HASVMYRGFGNNATANQY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 125546 seqs, 13197846 residues

125546

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Database :

Pending Patents AA.New.*
1: /cgn2_6/prodata/1/paa/PCT_NEW_COMB.pep.*
2: /cgn2_6/prodata/1/paa/US06_NEW_COMB.pep.*
3: /cgn2_6/prodata/1/paa/US07_NEW_COMB.pep.*
4: /cgn2_6/prodata/1/paa/US08_NEW_COMB.pep.*
5: /cgn2_6/prodata/1/paa/US09_NEW_COMB.pep.*
6: /cgn2_6/prodata/1/paa/US10_NEW_COMB.pep.*
7: /cgn2_6/prodata/1/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query #	Match Length	ID	Description
1	525	67.7	151	US-09-741-873C-4	Sequence 4, Appl
2	447	57.7	131	US-09-741-873C-2	Sequence 2, Appl
3	92	11.9	1327	PCT-US04-02338-49	Sequence 49, Appl
4	85.5	11.0	1048	PCT-US04-05654-2086	Sequence 2086, Ap
5	78.5	10.1	443	US-10-100-683-7608	Sequence 7608, Ap
6	77.5	10.0	318	US-10-417-884A-5818	Sequence 5818, Ap
7	77	9.9	125	US-10-767-701-58970	Sequence 58970, A
8	76.5	9.9	234	US-10-767-701-45603	Sequence 45603, A
9	76.5	9.9	386	US-10-100-683-10326	Sequence 10326, A
10	76.5	9.9	386	US-10-100-683-10327	Sequence 10327, A
11	76.5	9.9	386	US-10-093-037A-63	Sequence 63, Appl
12	76	9.8	250	US-10-779-461-40	Sequence 40, Appl
13	76	9.8	382	US-10-771-241-299	Sequence 299, Appl
14	75	9.7	191	US-10-767-701-47075	Sequence 47075, A
15	75	9.7	424	US-10-045-674A-591	Sequence 591, Appl
16	75	9.7	533	US-10-045-674A-5927	Sequence 527, Appl
17	74.5	9.6	392	US-10-417-884A-4840	Sequence 4840, Ap
18	74	9.5	485	PCT-US04-05654-1844	Sequence 1844, Ap
19	73	9.4	205	US-10-767-701-40104	Sequence 40104, A
20	73	9.4	246	US-10-779-461-41	Sequence 41, Appl
21	73	9.4	321	PCT-US04-05654-590	Sequence 590, Appl
22	72	9.3	434	US-10-045-674A-594	Sequence 594, Appl
23	72	9.3	1871	PCT-US03-02038-26	Sequence 26, Appl
24	71.5	9.2	179	US-10-767-701-426596	Sequence 42656, A
25	71.5	9.2	1160	US-10-603-150-2	Sequence 2, Appl
26	71	9.2	364	PCT-US04-05654-1756	Sequence 1756, Ap

27	71	9.2	447	US-10-767-701-45219	Sequence 45219, A
28	70.5	9.1	226	US-10-767-701-51685	Sequence 51685, A
29	70.5	9.1	251	US-10-767-701-32417	Sequence 32417, A
30	70.5	9.1	273	US-10-627-556-392	Sequence 392, Appl
31	70.5	9.1	273	US-10-627-556-396	Sequence 396, Appl
32	70.5	9.1	351	PCT-US04-05654-1755	Sequence 1755, Ap
33	70.5	9.1	505	US-10-627-556-400	Sequence 400, Appl
34	70.5	9.1	505	US-10-627-556-402	Sequence 402, Appl
35	70.5	9.1	521	US-10-627-556-398	Sequence 398, Appl
36	70	9.0	175	US-10-767-701-35046	Sequence 35046, A
37	70	9.0	199	US-10-767-701-51583	Sequence 51583, A
38	70	9.0	245	US-10-779-461-59	Sequence 59, Appl
39	70	9.0	348	US-10-004-115B-34	Sequence 34, Appl
40	70	9.0	624	PCT-US04-05654-1466	Sequence 1466, Ap
41	69	8.9	250	US-10-779-461-6	Sequence 6, Appl
42	69	8.9	616	US-10-467-243-1	Sequence 14, Appl
43	68.5	8.8	430	US-10-451-467A-314	Sequence 314, Appl
44	68	8.8	175	US-10-767-701-59665	Sequence 59665, A
45	68	8.8	251	US-10-779-461-30	Sequence 30, Appl

ALIGNMENTS

RESULT 1
US-09-741-873C-4
; Sequence 4, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873C-4

Query Match 67.7%; Score 525; DB 5; Length 151;

Best Local Similarity 68.2%; Pred. No. 36-40; Matches 103; Conservative 21; Mismatches 27; Indels 0; Gaps 0;

QY	1	MLKLKVAFAIVGSGSALAIVGPWGCGGNGGNSGSPSTLSIYYGSAANALAIQ	60
DB	1	MLKLKVAFAIVGSGSALAIVGPWGCGGNGGNSGSPSTLSIYYGSAANALAIQ	60
QY	61	SPARKSETTITGSGGNGADVGAGDNGSTIELTONGFRNNTITDQNAKNSDITVGOYQ	120
DB	61	TPARNSDLTTHGGGNGADVGQSSDDSDIDLQRFNGSATLDDQNGKNSRTVQFGC	120
QY	121	LVTRVTHEMASVYVRQFGNNATANQY	151
DB	121	GNGAVDQTAASSSVTVVQFGNNATANQY	151

RESULT 2

```

US-09-741-873C-2
; Sequence 2, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09741, 873C
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978, 878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347, 189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789, 437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970, 846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187, 865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318, 519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873C-2

```

Query Match 57.7%; Score 447; DB 5; Length 131;
 Best Local Similarity 64.9%; Pred. No. 2,3e-33;

Matches 85; Conservative 20; Mismatches 26; Indels 0; Gaps 0;

```

QY 21 GVFPWGGGNGGNGSGSPDSTLSIYQGSANALALQSDARKSETTITQSGYGAD 80
DB 1 GVFPYGGGNGGNGSGSPSEINTYQYGNSALALQTDANSIDITITQGGGNGAD 60
QY 81 VGGADNSTIELTQNGFNATIDQWNAKNSDITVGVQDQVTVTHEMAHSAVMROY 140
DB 61 VGGSDSDSIDITQRFNSATIDQWNAKNSDITVGVQDQVTVTHEMAHSAVMROY 120
QY 141 GFGNNAITANQY 151
DB 121 GFGNNAITANQY 131

```

RESULT 3

```

PCT-US04-02338-49
; Sequence 49, Application PC/TUS0402338
; GENERAL INFORMATION:
; APPLICANT: EXELIXIS, INC.
; TITLE OF INVENTION: MAPKs AS MODIFIERS OF THE APC AND AXIN PATHWAYS AND METHODS OF
; FILE REFERENCE: EX04-003C-PC
; CURRENT APPLICATION NUMBER: PCT/US04/02338
; PRIOR FILING DATE: 2004-01-28
; PRIOR APPLICATION NUMBER: US60/443,484
; PRIOR FILING DATE: 2003-01-29
; PRIOR APPLICATION NUMBER: US60/447,358
; PRIOR FILING DATE: 2003-02-11
; PRIOR APPLICATION NUMBER: US60/461,789
; PRIOR FILING DATE: 2003-04-10
; PRIOR APPLICATION NUMBER: US60/470,684
; PRIOR FILING DATE: 2003-05-14
; PRIOR APPLICATION NUMBER: US60/479,650
; PRIOR FILING DATE: 2003-06-19
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 49
; LENGTH: 1327

```

```

; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US04-02338-49

```

Query Match 11.9%; Score 92; DB 1; Length 1327;
 Best Local Similarity 30.4%; Pred. No. 1.9;
 Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

```

QY 6 VAAFAAI-VVSGALAGVQWGGGNGSGSPDSTLSIYQGSANALALQSDAR 64
DB 99 VAAFPVPAVTSAGVAPNPAQSGNSPSSSSPTSS-SSSPSSPSSSLAESPEAA 157
QY 65 KSETTIT---QSGYNGADVGGADNSTIELTQNG--FRNNATIDQWNAKNSDI 113
DB 158 GVSSTAPLPGGAAGPGTGVPAVSGALRELLACRNDGDAVRKLVDAAVNAKDM 212

```

RESULT 4

```

PCT-US04-05654-2086
; Sequence 2086, Application PC/TUS0405654
; GENERAL INFORMATION:
; APPLICANT: Sherman, Bradley K
; APPLICANT: Riechmann, Jose Luis
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Heard, Jacqueline E
; APPLICANT: Haake, Volker
; APPLICANT: Creelman, Robert A
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Adam, Luc J
; APPLICANT: Reuber, T. Lynne
; APPLICANT: Dubell, III, Arnold N
; APPLICANT: Pineda, Omaira
; APPLICANT: Repetti, Peter
; APPLICANT: Century, Karen
; APPLICANT: Guterson, Neal
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Brown, Pierre E
; APPLICANT: Kumamoto, Roderick W
; APPLICANT: Pilgrim, Marsha L
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES IN PLANTS
; FILE REFERENCE: MBI-0047 PCT
; CURRENT APPLICATION NUMBER: PCT/US04/05654
; PRIOR FILING DATE: 2004-03-03
; PRIOR APPLICATION NUMBER: 10/374,780
; PRIOR FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: 10/675,852
; PRIOR FILING DATE: 2003-09-30
; NUMBER OF SEQ ID NOS: 2950
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2086
; LENGTH: 1048
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: G1152 Paralogous to G1146
PCT-US04-05654-2086

```

Query Match 11.0%; Score 85.5; DB 1; Length 1048;
 Best Local Similarity 25.7%; Pred. No. 5.4;
 Matches 36; Conservative 17; Mismatches 60; Indels 27; Gaps 3;

```

QY 27 GGGGNGGNGSGSPDSTLSI-----YQGSANALALQSDARKSETTITQSGYG 76
DB 95 GGYGGGNGGSGSPDPQGVSPBELHQTSPYQAVSSQPTLSVSPYQVEPTVLAQOF- 153
QY 77 NGADVGGADNSTIE-----LTQNGFNATIDQWNAKNSDITVGVQYDQ 120
DB 154 EQLSVEQASQALQPIRSSKAFKFPMPKQSGSKCIIVKANHFAPALPRDLDHHYDV 213
QY 121 LVTRVTVTHEMAHSAVMROY 140
DB 214 TITPEVTSRGVNAVMKQLV 233

```

RESULT 5
US-10-683-7608
Sequence 7608, Application US/10100683
GENERAL INFORMATION:
APPLICANT: Rosen, et al.
TITLE OF INVENTION: Human Secreted Proteins
FILE REFERENCE: PS900
CURRENT APPLICATION NUMBER: US/10/100,683
CURRENT FILING DATE: 2002-03-19
PRIOR APPLICATION NUMBER: US 60/040,162
PRIOR FILING DATE: 1997-03-07
PRIOR APPLICATION NUMBER: US 60/043,576
PRIOR FILING DATE: 1997-04-11
PRIOR APPLICATION NUMBER: US 60/047,601
PRIOR FILING DATE: 1997-05-23
PRIOR APPLICATION NUMBER: US 60/056,845
PRIOR FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: US 60/043,580
PRIOR FILING DATE: 1997-04-11
PRIOR APPLICATION NUMBER: US 60/047,599
PRIOR FILING DATE: 1997-05-23
PRIOR APPLICATION NUMBER: US 60/056,664
PRIOR FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: US 60/043,314
PRIOR FILING DATE: 1997-04-11
PRIOR APPLICATION NUMBER: US 60/047,632
PRIOR FILING DATE: 1997-05-23
PRIOR APPLICATION NUMBER: US 60/056,892
PRIOR FILING DATE: 1997-08-22
Remaining Prior Application data removed - See file wrapper or PAM.
NUMBER OF SEQ ID NOS: 13468
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO: 7608
LENGTH: 443
TYPE: PRT
ORGANISM: Homo sapiens
US-10-683-7608

Query Match 10.1%; Score 78.5; DB 6; Length 443;
Best Local Similarity 25.0%; Pred. No. 7.9;
Matches 31; Conservative 17; Mismatches 43; Indels 33; Gaps 6;

QY 21 GVVPQ---WGCGNNHNGGNSGSDSTLSIYQ---YGSANAALALQSDARKSETTITQSGY 75
DB 185 GMPQAPMGCGG--NGPPNFGTNTQGAVALPGYGSVA-----SNQBGCTNPPSSG 237
QY 76 GNGAD---VQGGADNSTIELTONGFRNNATI-----DQNNAKNSDIT 114
DB 238 GGGSSNSGGGSGSSGSGSNGDNNNGSSGSGGSGGSRGDSGSSSNNGSSSTGSS 297
QY 115 VGOY 118
DB 298 SGNH 301

RESULT 6
US-10-417-884A-5818
Sequence 5818, Application US/10417884A
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESSER: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02154
COMPUTER READABLE FORM:

MEDIUM TYPE: CD/ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/417,884A
FILING DATE: 17-Apr-2003
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532
FILING DATE: 30-Jun-1998
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 5818:
SEQUENCE CHARACTERISTICS:
LENGTH: 318 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (8) LOCATION 1...318
SEQUENCE DESCRIPTION: SEQ ID NO: 5818:
US-10-417-884A-5818

Query Match 10.0%; Score 77.5; DB 6; Length 318;
Best Local Similarity 25.9%; Pred. No. 6.5;
Matches 30; Conservative 14; Mismatches 43; Indels 29; Gaps 4;

QY 6 VAAPAIWSSGALAGVWPQWGGGNNHNGGNSGSDSTLSIYQGSANAALALQSDARK 65
DB 19 VAFSAVLVAGC-----GADNNGSSNSSESSST-----AQSSTAK 53
QY 66 SETTITQSGYGVGCAVQGGADNSTIELTONGFRNNATIDQNNAKNSDITVGOYDOL 121
DB 54 S-TTSSASAVVAGSLDQCTVLEEKYNSNGYR---AVFEMVYKQKITESTKYDV 105

RESULT 7
US-10-767-701-58970
Sequence 58970, Application US/10767701
GENERAL INFORMATION:
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
FILE REFERENCE: 38-21(53535)B
CURRENT APPLICATION NUMBER: US/10/767,701
CURRENT FILING DATE: 2004-01-29
NUMBER OF SEQ ID NOS: 63128
SEQ ID NO: 58970
LENGTH: 125
TYPE: PRT
ORGANISM: Sorghum bicolor
FEATURE:
NAME/KEY: unsure
LOCATION: (1)..(125)
OTHER INFORMATION: unsure at all Xaa locations
FEATURE:
OTHER INFORMATION: Clone ID: 6858533.pep
US-10-767-701-58970

Query Match 9.9%; Score 77; DB 6; Length 125;
 Best Local Similarity 25.9%; Pred. No. 2.3;
 Matches 30; Conservative 11; Mismatches 49; Indels 26; Gaps 4;

QY 24 PGMGGGNNH--NGGSSGPDSTLSIYQGSANALALQSDARKSETTTTOSGYGNG-- 78
 DB 14 POGYGNNGYNGGGYNSPP-----AQGYGSVEMAPLANHA-----GFGGQDDP 59
 QY 79 -----ADVGGANSTIELTONGFRNNATIDQNNAKNSDITVQYDQLYTRVYT 127
 DB 60 NALNECRDIDRGIDPVEQNLEQLRMIOQRITLDAXSSGSSAASRQDLSLSTDTIS 115

RESULT 8
 US-10-767-701-45603
 ; Sequence 45603, Application US/10767701
 ; GENERAL INFORMATION:
 ; APPLICANT: Kovalic, David K.
 ; APPLICANT: Zhou, Yihua
 ; APPLICANT: Cao, Yongwei
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
 ; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
 ; FILE REFERENCE: 38-21(5355)B
 ; CURRENT APPLICATION NUMBER: US/10/767,701
 ; NUMBER OF SEQ ID NOS: 2004-01-29
 ; SEQ ID NO 45603
 ; LENGTH: 234
 ; TYPE: PRT
 ; ORGANISM: Sorghum bicolor
 ; FEATURE:
 ; NAME/KEY: unsure
 ; LOCATION: (1)..(234)
 ; OTHER INFORMATION: unsure at all Xaa locations
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: SORBI-28MAY03-C82834_1.pep
 ; US-10-767-701-45603

Query Match 9.9%; Score 76.5; DB 6; Length 234;
 Best Local Similarity 23.7%; Pred. No. 5.4;
 Matches 37; Conservative 10; Mismatches 44; Indels 65; Gaps 6;

QY 1 MKLKVAFALIV--SGSALAGVPPW--GGGNNNGGN-----SSGPDSTL 45
 DB 1 MATTKALCPVILIGGNAARVARYVSAGCGGCGGGRGASRMGSGSGSGC 60
 QY 46 SIY-----OYGSANALALQSDARKSETTTTOSGYGNGADV 82
 DB 61 GXYEAGSGXAYAGCGGCGGCGGGOYIGSG-----SGYGSBGY 102
 QY 83 QGADNSTIELTONGFRNNATIDQNNAKNSDITVQY 118
 DB 103 QAGSGSGS-----NGAYAGGAGCGGCGGCGGOY 129

RESULT 9
 US-10-100-683-10326
 ; Sequence 10326, Application US/10100683
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen, et al.
 ; TITLE OF INVENTION: Human Secreted Proteins
 ; FILE REFERENCE: PS900
 ; CURRENT APPLICATION NUMBER: US/10/100,683
 ; PRIOR FILING DATE: 2002-03-19
 ; PRIOR APPLICATION NUMBER: US 60/040,162
 ; PRIOR FILING DATE: 1997-03-07
 ; PRIOR APPLICATION NUMBER: US 60/043,576
 ; PRIOR FILING DATE: 1997-04-11
 ; PRIOR APPLICATION NUMBER: US 60/047,601
 ; PRIOR FILING DATE: 1997-05-23
 ; PRIOR APPLICATION NUMBER: US 60/056,845
 ; PRIOR FILING DATE: 1997-08-22

PRIOR APPLICATION NUMBER: US 60/043,580
 ; PRIOR FILING DATE: 1997-04-11
 ; PRIOR APPLICATION NUMBER: US 60/047,599
 ; PRIOR FILING DATE: 1997-05-23
 ; PRIOR APPLICATION NUMBER: US 60/056,664
 ; PRIOR FILING DATE: 1997-08-22
 ; PRIOR APPLICATION NUMBER: US 60/043,314
 ; PRIOR FILING DATE: 1997-04-11
 ; PRIOR APPLICATION NUMBER: US 60/047,632
 ; PRIOR FILING DATE: 1997-05-23
 ; PRIOR APPLICATION NUMBER: US 60/056,892
 ; PRIOR FILING DATE: 1997-08-22
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 13468
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 10326
 ; LENGTH: 386
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-10-100-683-10326

Query Match 9.9%; Score 76.5; DB 6; Length 386;
 Best Local Similarity 31.1%; Pred. No. 10;
 Matches 28; Conservative 13; Mismatches 34; Indels 15; Gaps 5;

QY 21 GVPQ---WGGGNNHGGSSGPDSTLSIY--YGSANALALQSDARKSETTTTOSGY 75
 DB 185 GAMPQAPWQGG--NGGPNFNTVQGVAPGYGSVRA-----SNONEGCTNPPPSGS 237
 QY 76 GNGAD---VGGADNSTIELTONGFRNNAT 102
 DB 238 GGGSSNSGGSGSGSSGSSGSSGSGNNGS 267

RESULT 10
 US-10-100-683-10327
 ; Sequence 10327, Application US/10100683
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen, et al.
 ; TITLE OF INVENTION: Human Secreted Proteins
 ; FILE REFERENCE: PS900
 ; CURRENT APPLICATION NUMBER: US/10/100,683
 ; PRIOR FILING DATE: 2002-03-19
 ; PRIOR APPLICATION NUMBER: US 60/040,162
 ; PRIOR FILING DATE: 1997-03-07
 ; PRIOR APPLICATION NUMBER: US 60/043,576
 ; PRIOR FILING DATE: 1997-04-11
 ; PRIOR APPLICATION NUMBER: US 60/047,601
 ; PRIOR FILING DATE: 1997-05-23
 ; PRIOR APPLICATION NUMBER: US 60/056,845
 ; PRIOR FILING DATE: 1997-08-22
 ; PRIOR APPLICATION NUMBER: US 60/043,580
 ; PRIOR FILING DATE: 1997-04-11
 ; PRIOR APPLICATION NUMBER: US 60/047,599
 ; PRIOR FILING DATE: 1997-05-23
 ; PRIOR APPLICATION NUMBER: US 60/056,664
 ; PRIOR FILING DATE: 1997-08-22
 ; PRIOR APPLICATION NUMBER: US 60/043,314
 ; PRIOR FILING DATE: 1997-04-11
 ; PRIOR APPLICATION NUMBER: US 60/047,632
 ; PRIOR FILING DATE: 1997-05-23
 ; PRIOR APPLICATION NUMBER: US 60/056,892
 ; PRIOR FILING DATE: 1997-08-22
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 13468
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 10327
 ; LENGTH: 386
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-10-100-683-10327

Query Match 9.9%; Score 76.5; DB 6; Length 386;


```

; LENGTH: 191
;
; TYPE: PR1
;
; ORGANISM: Sorghum bicolor
;
; FEATURE:
;
; OTHER INFORMATION: Clone ID: SORBI-28MA03-C24509_1.pep
US-10-767-701-47075

```

Query Match	9.7%;	Score 75;	DB 6;	Length 191;
Best Local Similarity	28.4%;	Pred. No. 5.8;		
Matches 33;	Conservative 9;	Mismatches 46;	Indels 28;	Gaps 4;

```

QY 7 AAFPAIVSSSALACVPPQWGGGGNNHGGNSSGPPSTLSIQYGSANAALLQSDARKS 66
      ||| ||| ||| :| :| :| :|
DQ 85 AAAAAAGAAAGSAAAGS--GSAATGSGAGAGAGSTTLA-----SSAGAA----- 128

```

QY 67 ETTITSGYGNADVGGADNSTIELTQNGFRNNATIDQWAKNSDITVGQYDQLV 122
 DQ 129 -----SSGTAGASASAGAGAA-----GTTAGAGVSTTGASSVATTGVWITSAY 172

RESULT 15

```

; Sequence 591, Application US/10045674A
; US-10-045-674A-591
; GENERAL INFORMATION:
; APPLICANT: LADNER, ROBERT C.
; APPLICANT: COHEN, EDWARD H.
; APPLICANT: NASTRI, HORACIO G.
; APPLICANT: ROOKEY, KRISTIN L.
; APPLICANT: HOET, RENE
; APPLICANT: HOOGENBOOM, HENDRICUS R. J. M.
; TITLE OF INVENTION: NOVEL METHODS OF CONSTRUCTING LIBRARIES COMPRISING
; TITLE OF INVENTION: DISPLAYED AND/OR EXPRESSED MEMBERS OF A DIVERSE FAMILY
; TITLE OF INVENTION: OF PEPTIDES, POLYPEPTIDES OR PROTEINS AND THE NOVEL
; FILE REFERENCE: DYAX/002 CIP2
; CURRENT APPLICATION NUMBER: US/10/045,674A
; CURRENT FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: 06/198,069
; PRIOR FILING DATE: 2000-04-17
; PRIOR APPLICATION NUMBER: 09/637,306
; PRIOR FILING DATE: 2001-04-17
; NUMBER OF SEQ ID NOS: 635
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 591
; LENGTH: 424
; TYPE: PRT
; ORGANISM: Unknown Organism
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: M13 protein
; US-10-045-674A-591

```

Query Match	9.7%	Score 75;	DB 6;	Length 424;
Best Local Similarity	30.3%;	Pred. No. 15;		
Matches 37; Conservative	8;	Mismatches 35;	Indels 42;	Gaps 8;

QY 15 SCSALAGVWPQMG--GGNNNGGSSGPDSTLSIQ-YGSANAAL-----ALQSDA 63
 || : || ||| : ||| : |||
 Db 243 SGGSGEGGSGEGGSGEGGSGGSGDFFYEKMANANKGANTENADENALQSDA 302

QY 64 R-KSETITIQ-----SGVNG-----ADVGGADNSTIELTQNG 96
 : : : | | |
Dd 303 KGRLDVAVDYGAAIDGFLGDVSGLANGANGATGDFAGSNQMAQVEVG--DNSPL--MNN 358

Qy	97	FR	98
Db	359	FR	360

Search completed: March 11, 2004, 19:14:15
Job time : 5.3 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using ew model

Run on: March 11, 2004, 18:33:51 ; Search time 171.3 Seconds
(without alignments)
860.386 Million cell updates/sec

Title: US-09-543-407-16

Perfect score: 757
Sequence: 1 MLKLKVAAPAAIWSGSALA.....DSVMVROYGFNNATFANQY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 6019581 seqs, 976053577 residues

Total number of hits satisfying chosen parameters: 6019581

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Pending Patents, AA, Main:*

1: /cgn2_6/ptodata/2/paa/US06_COMB.pep.*
2: /cgn2_6/ptodata/2/paa/US06_COMB.pep.*
3: /cgn2_6/ptodata/2/paa/US07_COMB.pep.*
4: /cgn2_6/ptodata/2/paa/US08_COMB.pep.*
5: /cgn2_6/ptodata/2/paa/US08_COMB.pep.*
6: /cgn2_6/ptodata/2/paa/US08_COMB.pep.*
7: /cgn2_6/ptodata/2/paa/US08_COMB.pep.*
8: /cgn2_6/ptodata/2/paa/US08_COMB.pep.*
9: /cgn2_6/ptodata/2/paa/US08_COMB.pep.*
10: /cgn2_6/ptodata/2/paa/US08_COMB.pep.*
11: /cgn2_6/ptodata/2/paa/US08_COMB.pep.*
12: /cgn2_6/ptodata/2/paa/US08_COMB.pep.*
13: /cgn2_6/ptodata/2/paa/US08_COMB.pep.*
14: /cgn2_6/ptodata/2/paa/US09_COMB.pep.*
15: /cgn2_6/ptodata/2/paa/US09_COMB.pep.*
16: /cgn2_6/ptodata/2/paa/US09_COMB.pep.*
17: /cgn2_6/ptodata/2/paa/US09_COMB.pep.*
18: /cgn2_6/ptodata/2/paa/US09_COMB.pep.*
19: /cgn2_6/ptodata/2/paa/US09_COMB.pep.*
20: /cgn2_6/ptodata/2/paa/US09_COMB.pep.*
21: /cgn2_6/ptodata/2/paa/US09_COMB.pep.*
22: /cgn2_6/ptodata/2/paa/US09_COMB.pep.*
23: /cgn2_6/ptodata/2/paa/US09_COMB.pep.*
24: /cgn2_6/ptodata/2/paa/US09_COMB.pep.*
25: /cgn2_6/ptodata/2/paa/US09_COMB.pep.*
26: /cgn2_6/ptodata/2/paa/US10_COMB.pep.*
27: /cgn2_6/ptodata/2/paa/US10_COMB.pep.*
28: /cgn2_6/ptodata/2/paa/US10_COMB.pep.*
29: /cgn2_6/ptodata/2/paa/US10_COMB.pep.*
30: /cgn2_6/ptodata/2/paa/US10_COMB.pep.*
31: /cgn2_6/ptodata/2/paa/US10_COMB.pep.*
32: /cgn2_6/ptodata/2/paa/US10_COMB.pep.*
33: /cgn2_6/ptodata/2/paa/US10_COMB.pep.*

Prod. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
------------	-------	--------------------	----	-------------

1	757	100.0	151	19	US-09-543-407-16	Sequence 16, Appl
2	658	86.9	151	19	US-09-543-407-5	Sequence 5, Appl
3	653	86.3	151	6	US-08-233-642A-57	Sequence 57, Appl
4	618	81.6	151	19	US-09-543-407-18	Sequence 18, Appl
5	580	76.6	151	19	US-09-543-407-12	Sequence 12, Appl
6	579	76.5	131	19	US-09-543-407-31	Sequence 31, Appl
7	578	76.4	151	19	US-09-543-407-14	Sequence 14, Appl
8	577	76.2	151	19	US-09-543-407-26	Sequence 26, Appl
9	575	76.0	151	19	US-09-543-407-24	Sequence 24, Appl
10	574	75.8	151	19	US-09-543-407-20	Sequence 20, Appl
11	568	75.0	151	19	US-09-543-407-28	Sequence 28, Appl
12	567	74.9	151	19	US-09-543-407-22	Sequence 22, Appl
13	566	74.8	151	19	US-09-543-407-30	Sequence 30, Appl
14	554	73.2	109	19	US-09-543-407-34	Sequence 34, Appl
15	487	64.3	120	6	US-08-233-642A-55	Sequence 55, Appl
16	484	63.9	151	19	US-09-543-407-7	Sequence 7, Appl
17	481	63.5	151	13	US-08-978-878-4	Sequence 4, Appl
18	481	63.5	151	21	US-09-741-873B-4	Sequence 4, Appl
19	479	63.3	151	33	US-60-352-946-2	Sequence 2, Appl
20	479	63.3	151	33	US-60-444-371-2	Sequence 2, Appl
21	468	61.8	158	16	US-09-252-691C-5834	Sequence 5834, Ap
22	468	61.8	158	16	US-09-252-691C-5834	Sequence 5834, Ap
23	468	61.8	158	30	US-10-417-886-5834	Sequence 5834, Ap
24	414	54.7	131	13	US-08-978-878-2	Sequence 2, Appl
25	414	54.7	131	21	US-09-741-873B-2	Sequence 2, Appl
26	401	53.0	109	19	US-09-543-407-35	Sequence 35, Appl
27	347	45.8	68	19	US-09-543-407-37	Sequence 37, Appl
28	250.5	33.1	70	19	US-09-543-407-32	Sequence 32, Appl
29	237	31.3	48	19	US-09-543-407-39	Sequence 39, Appl
30	122.5	16.2	186	16	US-09-252-691C-5833	Sequence 5833, Ap
31	122.5	16.2	186	16	US-09-252-691C-5833	Sequence 5833, Ap
32	122.5	16.2	186	30	US-10-417-886-5833	Sequence 5833, Ap
33	118.5	15.7	151	19	US-09-543-407-8	Sequence 8, Appl
34	113.5	15.0	445	29	US-10-369-493-20638	Sequence 20638, A
35	113.5	15.0	445	33	US-60-360-039-20638	Sequence 20638, A
36	111.5	14.7	151	19	US-09-543-407-6	Sequence 6, Appl
37	101.5	13.4	590	3	US-07-965-678A-10	Sequence 10, Appl
38	101.5	13.4	590	22	US-09-791-537-29229	Sequence 29229, A
39	101.5	13.4	590	22	US-09-791-537-111372	Sequence 111372, A
40	101	13.3	599	22	US-09-791-537-29228	Sequence 29228, A
41	101	13.3	599	22	US-09-791-537-111371	Sequence 111371, A
42	99.5	13.1	582	22	US-09-791-537-94843	Sequence 94843, A
43	99.5	13.1	646	22	US-09-791-537-30765	Sequence 30765, A
44	99.5	13.1	646	22	US-09-791-537-32862	Sequence 32862, A
45	98	12.9	145	21	US-09-739-449-8854	Sequence 8854, Ap

ALIGNMENTS

RESULT 1
US-09-543-407-16
Sequence 16, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543,407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 16
LENGTH: 151
TYPE: PRT
ORGANISM: Artificial sequence
FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
OTHER INFORMATION: Sequence containing the replacement fragment
encoding PT3 from GP63 of Leishmania major.

US-09-543-407-16

Query Match

Best Local Similarity 100.0%; Score 757; DB 19; Length 151;
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLKVAFALVIVSGSALAGVVDLVTRVTHMAHAGSDSTLSIYOGSANAALALQ 60
DB 1 MKLKVAFALVIVSGSALAGVVDLVTRVTHMAHAGSDSTLSIYOGSANAALALQ 60
QY 61 SDARKSETTITQSGYGADVGAGADNSTIELTONGFRNNATIDQWNAKNSDITVGOYGG 120
DB 61 SDARKSETTITQSGYGADVGAGADNSTIELTONGFRNNATIDQWNAKNSDITVGOYGG 120
QY 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151
DB 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151

RESULT 2

US-09-543-407-5
Sequence 5, Application US/09543407
GENERAL INFORMATION:

APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543,407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5
LENGTH: 151
TYPE: PRT
ORGANISM: Salmonella enteritidis
US-09-543-407-5

Query Match 86.9%; Score 658; DB 19; Length 151;
Best Local Similarity 90.1%; Pred. No. 4.5e-65;

Matches 136; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

QY 1 MKLKVAFALVIVSGSALAGVVDLVTRVTHMAHAGSDSTLSIYOGSANAALALQ 60
DB 1 MKLKVAFALVIVSGSALAGVVDLVTRVTHMAHAGSDSTLSIYOGSANAALALQ 60
QY 61 SDARKSETTITQSGYGADVGAGADNSTIELTONGFRNNATIDQWNAKNSDITVGOYGG 120
DB 61 SDARKSETTITQSGYGADVGAGADNSTIELTONGFRNNATIDQWNAKNSDITVGOYGG 120
QY 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151
DB 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151

RESULT 3

US-08-233-642A-57
Sequence 57, Application US/08233642A
GENERAL INFORMATION:

APPLICANT: Kay, William W.
APPLICANT: Collinson, S. Karen
APPLICANT: Clouthier, Sharon C.
APPLICANT: Doran, James L.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
NUMBER OF SEQUENCES: 58
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed and Berry
STREET: 6300 Columbia Center, 701 Fifth Avenue
City: Seattle

STATE: Washington

COUNTRY: U.S.A.

ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/233,642A

FILING DATE: 26-Apr-1994

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: King, Joshua

REGISTRATION NUMBER: 35,570

REFERENCE/DOCKET NUMBER: 920043.403C3

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 57:

SEQUENCE CHARACTERISTICS:

LENGTH: 151 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-233-642A-57

Query Match

Best Local Similarity 86.3%; Score 653; DB 6; Length 151;
Matches 135; Conservative 3; Mismatches 13; Indels 0; Gaps 0;

QY 1 MKLKVAFALVIVSGSALAGVVDLVTRVTHMAHAGSDSTLSIYOGSANAALALQ 60
DB 1 MKLKVAFALVIVSGSALAGVVDLVTRVTHMAHAGSDSTLSIYOGSANAALALQ 60
QY 61 SDARKSETTITQSGYGADVGAGADNSTIELTONGFRNNATIDQWNAKNSDITVGOYGG 120
DB 61 SDARKSETTITQSGYGADVGAGADNSTIELTONGFRNNATIDQWNAKNSDITVGOYGG 120
QY 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151
DB 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151

RESULT 4

US-09-543-407-18
Sequence 18, Application US/09543407
GENERAL INFORMATION:

APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543,407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 18
LENGTH: 151
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b atga
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding P73 from GP63 of Leishmania major.
US-09-543-407-18

Query Match

Best Local Similarity 81.6%; Score 618; DB 19; Length 151;
Matches 131; Conservative 0; Mismatches 0; Indels 40; Gaps 2;

```

QY      1 MLLKVAAPAAIVVSGSALAGV-----VDQVTRVVTHEMAHAG 40
      1 MLLKVAAPAAIVVSGSALAGVPPWGGGNNHGGNSGSPDQVTRVVTHEMAH--- 57
DB      41 POSTLSIYQGSANALALQSPARKSETTITGSGYNGADVGGADNSTIELTONGFRNN 100
QY      58 -----ALQSPARKSETTITGSGYNGADVGGADNSTIELTONGFRNN 100
DB      101 ATIDQNAKNSDITVQYGGNNALVNOTASDSSVWVROVGFNNATANQY 151
      101 ATIDQNAKNSDITVQYGGNNALVNOTASDSSVWVROVGFNNATANQY 151
DB      101 ATIDQNAKNSDITVQYGGNNALVNOTASDSSVWVROVGFNNATANQY 151

RESULT 5
US-09-543-407-12
/ Sequence 12, Application US/09543407
/ GENERAL INFORMATION:
/ APPLICANT: White, Aaron P.
/ APPLICANT: Doran, James L.
/ APPLICANT: Collinson, S. Karen
/ APPLICANT: Kay, William W.
/ TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
/ FILE REFERENCE: 920043.406
/ CURRENT APPLICATION NUMBER: US/09/543,407
/ CURRENT FILING DATE: 2000-04-05
/ NUMBER OF SEQ ID NOS: 59
/ SOFTWARE: FaastSeq for Windows Version 4.0
/ SEQ ID NO 12
/ LENGTH: 151
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
/ OTHER INFORMATION: sequence containing the replacement fragment
/ OTHER INFORMATION: encoding PT3 from GPe3 of Leishmania major.
US-09-543-407-12

Query Match
Best Local Similarity 76.6%; Score 580; DB 19; Length 151;
Matches 121; Conservative 7; Mismatches 23; Indels 0; Gaps 0;

QY      1 MLLKVAAPAAIVVSGSALAGVVDQVTRVVTHEMAHAGSPDSTLSIYQGSANALALQ 60
      1 MLLKVAAPAAIVVSGSALAGVPPWGGGNNHGGNSGSPDSTLSIYQGSANALALQ 60
DB      61 SPARKSETTITGSGYNGADVGGADNSTIELTONGFRNNATIDQNAKNSDITVQYGG 120
      61 SPARKSETTITGSGYNGADVGGADNSTIELTONGFRNNATIDQNAKNSDITVQYGG 120
DB      61 SPARKSETTITGSGYNGADVGGADNSTIELTONGFRNNATIDQNAKNSDITVQYGG 120
QY      121 NNAALVNOTASDSSVWVROVGFNNATANQY 151
      121 NNAALVNOTASDSSVWVROVGFNNATANQY 151
DB      121 NNAALVNOTASDSSVWVROVGFNNATANQY 151

RESULT 6
US-09-543-407-31
/ Sequence 31, Application US/09543407
/ GENERAL INFORMATION:
/ APPLICANT: White, Aaron P.
/ APPLICANT: Doran, James L.
/ APPLICANT: Collinson, S. Karen
/ APPLICANT: Kay, William W.
/ TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
/ FILE REFERENCE: 920043.406
/ CURRENT APPLICATION NUMBER: US/09/543,407
/ CURRENT FILING DATE: 2000-04-05
/ NUMBER OF SEQ ID NOS: 59
/ SOFTWARE: FaastSeq for Windows Version 4.0
/ SEQ ID NO 31
/ LENGTH: 131

```

```

/ TYPE: PRT
/ ORGANISM: Salmonella enteritidis
US-09-543-407-31

Query Match
Best Local Similarity 76.5%; Score 579; DB 19; Length 131;
Matches 113; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      37 HASGPDSTLSIYQGSANALALQSPARKSETTITGSGYNGADVGGADNSTIELTONG 96
      37 HASGPDSTLSIYQGSANALALQSPARKSETTITGSGYNGADVGGADNSTIELTONG 96
DB      17 NSGSPDSTLSIYQGSANALALQSPARKSETTITGSGYNGADVGGADNSTIELTONG 76
QY      97 FRNNATIDQNAKNSDITVQYGGNNALVNOTASDSSVWVROVGFNNATANQY 151
      97 FRNNATIDQNAKNSDITVQYGGNNALVNOTASDSSVWVROVGFNNATANQY 151
DB      77 FRNNATIDQNAKNSDITVQYGGNNALVNOTASDSSVWVROVGFNNATANQY 131

RESULT 7
US-09-543-407-14
/ Sequence 14, Application US/09543407
/ GENERAL INFORMATION:
/ APPLICANT: White, Aaron P.
/ APPLICANT: Doran, James L.
/ APPLICANT: Collinson, S. Karen
/ APPLICANT: Kay, William W.
/ TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
/ FILE REFERENCE: 920043.406
/ CURRENT APPLICATION NUMBER: US/09/543,407
/ CURRENT FILING DATE: 2000-04-05
/ NUMBER OF SEQ ID NOS: 59
/ SOFTWARE: FaastSeq for Windows Version 4.0
/ SEQ ID NO 14
/ LENGTH: 151
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
/ OTHER INFORMATION: sequence containing the replacement fragment
/ OTHER INFORMATION: encoding PT3 from GPe3 of Leishmania major.
US-09-543-407-14

Query Match
Best Local Similarity 76.4%; Score 578; DB 19; Length 151;
Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;

QY      1 MLLKVAAPAAIVVSGSALAGVVDQVTRVVTHEMAHAGSPDSTLSIYQGSANALALQ 60
      1 MLLKVAAPAAIVVSGSALAGVPPWGGGNNHGGNSGSPDSTLSIYQGSANALALQ 60
DB      61 SPARKSETTITGSGYNGADVGGADNSTIELTONGFRNNATIDQNAKNSDITVQYGG 120
      61 SPARKSETTITGSGYNGADVGGADNSTIELTONGFRNNATIDQNAKNSDITVQYGG 120
DB      61 SPARKSETTITGSGYNGADVGGADNSTIELTONGFRNNATIDQNAKNSDITVQYGG 120
QY      121 NNAALVNOTASDSSVWVROVGFNNATANQY 151
      121 NNAALVNOTASDSSVWVROVGFNNATANQY 151
DB      121 LTVTRVVTHEMAHAGVWVROVGFNNATANQY 151

RESULT 8
US-09-543-407-26
/ Sequence 26, Application US/09543407
/ GENERAL INFORMATION:
/ APPLICANT: White, Aaron P.
/ APPLICANT: Doran, James L.
/ APPLICANT: Collinson, S. Karen
/ APPLICANT: Kay, William W.
/ TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
/ FILE REFERENCE: 920043.406
/ CURRENT APPLICATION NUMBER: US/09/543,407
/ CURRENT FILING DATE: 2000-04-05
/ NUMBER OF SEQ ID NOS: 59

```

```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-26

```

```

Query Match
Best Local Similarity 76.2%; Score 577; DB 19; Length 151;
Matches 122; Conservative 5; Mismatches 24; Indels 0; Gaps 0;

```

```

QY 1 MKLKVAFPAIIVSGSALAGVYDQVTRVVTHEMAHSGPDSSTLIYOGSANAALALQ 60
DB 1 MKLKVAFPAIIVSGSALAGVYDQVTRVVTHEMAHSGPDSSTLIYOGSANAALALQ 60
QY 61 SDARKSETTITGSGYNGADVGGADNSTIELTONGFRNNATTIDQNAKNSDITVGYG 120
DB 61 LVTRVVTHEMAHAGYNGADVGGADNSTIELTONGFRNNATTIDQNAKNSDITVGYG 120

```

```

QY 121 NNAALVNOTASDSVWVRQVGFNNATANQY 151
DB 121 NNAALVNOTASDSVWVRQVGFNNATANQY 151

```

```

RESULT 9
US-09-543-407-24

```

```

; Sequence 24, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543.407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-24

```

```

Query Match
Best Local Similarity 76.0%; Score 575; DB 19; Length 151;
Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;

```

```

QY 1 MKLKVAFPAIIVSGSALAGVYDQVTRVVTHEMAHSGPDSSTLIYOGSANAALALQ 60
DB 1 MKLKVAFPAIIVSGSALAGVYDQVTRVVTHEMAHSGPDSSTLIYOGSANAALALQ 60
QY 61 SDARKSETTITGSGYNGADVGGADNSTIELTONGFRNNATTIDQNAKNSDITVGYG 120
DB 61 SDARKSETTITGSGYNGADVGGADNSTIELTONGFRNNATTIDQNAKNSDITVGYG 120

```

```

QY 121 NNAALVNOTASDSVWVRQVGFNNATANQY 151
DB 121 NNAALVNOTASDSVWVRQVGFNNATANQY 151

```

```

RESULT 10
US-09-543-407-20
; Sequence 20, Application US/09543407

```

```

; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543.407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-20

```

```

Query Match
Best Local Similarity 75.8%; Score 574; DB 19; Length 151;
Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;

```

```

QY 1 MKLKVAFPAIIVSGSALAGVYDQVTRVVTHEMAHSGPDSSTLIYOGSANAALALQ 60
DB 1 MKLKVAFPAIIVSGSALAGVYDQVTRVVTHEMAHSGPDSSTLIYOGSANAALALQ 60
QY 61 SDARKSETTITGSGYNGADVGGADNSTIELTONGFRNNATTIDQNAKNSDITVGYG 120
DB 61 SDARKSETTITGSGYNGADVGGADNSTIELTONGFRNNATTIDQNAKNSDITVGYG 120

```

```

QY 121 NNAALVNOTASDSVWVRQVGFNNATANQY 151
DB 121 NNAALVNOTASDSVWVRQVGFNNATANQY 151

```

```

RESULT 11
US-09-543-407-28

```

```

; Sequence 28, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543.407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-28

```

```

Query Match
Best Local Similarity 75.0%; Score 568; DB 19; Length 151;
Matches 121; Conservative 4; Mismatches 26; Indels 0; Gaps 0;

```

```

QY 1 MKLKVAFPAIIVSGSALAGVYDQVTRVVTHEMAHSGPDSSTLIYOGSANAALALQ 60
DB 1 MKLKVAFPAIIVSGSALAGVYDQVTRVVTHEMAHSGPDSSTLIYOGSANAALALQ 60
QY 61 SDARKSETTITGSGYNGADVGGADNSTIELTONGFRNNATTIDQNAKNSDITVGYG 120
DB 61 SDARKSETTITGSGYNGADVGGADNSTIELTONGFRNNATTIDQNAKNSDITVGYG 120

```

```

QY 121 NNAALVNOTASDSVWVRQVGFNNATANQY 151
DB 121 NNAALVNOTASDSVWVRQVGFNNATANQY 151

```

Db 61 SDARKSETTITGSGYNGADYDQVTRVVTHEMAHAFRNATIDQWNAKSDITVGOYGG 120
QY 121 NNAALVNOTASDSSVMRVQVFGNNATANOY 151
Db 121 NNAALVNOTASDSSVMRVQVFGNNATANOY 151

RESULT 12

US-09-543-407-22
; Sequence 22, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FaastSeq for Windows Version 4.0
; SEQ ID NO 22
; LENGTH: 151
; TYPE: PR1
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
US-09-543-407-22

Query Match 74.9%; Score 567; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 7.5e-55;
Matches 122; Conservative 5; Mismatches 24; Indels 0; Gaps 0;

QY 1 MLLKVAAPAAIVSGSALAGYDQVTRVVTHEMAHAGPSTLSIYOGSANAALAQ 60
Db 1 MLLKVAAPAAIVSGSALAGYDQVTRVVTHEMAHAGPSTLSIYOGSANAALAQ 60
QY 61 SDARKSETTITGSGYNGADYDQVTRVVTHEMAHAGPSTLSIYOGSANAALAQ 120
Db 61 SDARKSETTITGSGYNGADYDQVTRVVTHEMAHAGPSTLSIYOGSANAALAQ 120
QY 121 NNAALVNOTASDSSVMRVQVFGNNATANOY 151
Db 121 NNAALVNOTASDSSVMRVQVFGNNATANOY 151

RESULT 13

US-09-543-407-30
; Sequence 30, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FaastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 151
; TYPE: PR1
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GPe3 of Leishmania major.

US-09-543-407-30

Query Match 74.8%; Score 566; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 9.7e-55;
Matches 122; Conservative 5; Mismatches 24; Indels 0; Gaps 0;

QY 1 MLLKVAAPAAIVSGSALAGYDQVTRVVTHEMAHAGPSTLSIYOGSANAALAQ 60
Db 1 MLLKVAAPAAIVSGSALAGYDQVTRVVTHEMAHAGPSTLSIYOGSANAALAQ 60
QY 61 SDARKSETTITGSGYNGADYDQVTRVVTHEMAHAGPSTLSIYOGSANAALAQ 120
Db 61 SDARKSETTITGSGYNGADYDQVTRVVTHEMAHAGPSTLSIYOGSANAALAQ 120
QY 121 NNAALVNOTASDSSVMRVQVFGNNATANOY 151
Db 121 NNAALVNOTASDSSVMRVQVFGNNATANOY 151

RESULT 14

US-09-543-407-34
; Sequence 34, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FaastSeq for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 109
; TYPE: PR1
; ORGANISM: Salmonella enteritidis
US-09-543-407-34

Query Match 73.2%; Score 554; DB 19; Length 109;
Best Local Similarity 100.0%; Pred. No. 1.4e-53;
Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 43 STLSIYOGSANAALAQSDARKSETTITGSGYNGADYDQVTRVVTHEMAHAGPSTLSIYOGSANAALAQ 102
Db 1 STLSIYOGSANAALAQSDARKSETTITGSGYNGADYDQVTRVVTHEMAHAGPSTLSIYOGSANAALAQ 102
QY 103 IDQWNAKSDITVGOYGGNNAALVNOTASDSSVMRVQVFGNNATANOY 151
Db 61 IDQWNAKSDITVGOYGGNNAALVNOTASDSSVMRVQVFGNNATANOY 109

RESULT 15

US-08-233-642A-55
; Sequence 55, Application US/08233642A
; GENERAL INFORMATION:
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Cloutier, Sharon C.
; APPLICANT: Doran, James L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
; BASED VACCINES
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/233,642A
 FILING DATE: 26-APR-1994
 CLASSIFICATION: 424
 ATTORNEY/AGENT INFORMATION:
 NAME: King, Joshua
 REGISTRATION NUMBER: 35,570
 REFERENCE/DOCKET NUMBER: 920043,403C3
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 622-4900
 TELEFAX: (206) 682-6031
 TELEX: 3723836 SEDANBERRY
 INFORMATION FOR SEQ ID NO: 55:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 120 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-233-642A-55

Query Match 64.3%; Score 487; DB 6; Length 120;
 Best Local Similarity 97.9%; Pred. No. 5.2e-46;
 Matches 95; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 37 HASGPDSTLSIYOGSANAALALOSDARKSETTTTOSGYGNGADVGGADNSTIELTONG 96
 DB 16 NSSGPDSTLSIYOGSANAALALOSDARKSETTTTOSGYGNGADVGGADNSTIELTONG 75
 QY 97 FRNNATIDQWNAKNSDITVGOYGNNALVNOQTASDS 133
 DB 76 FRNNATIDQWNAKNSDITVGOYGNNALVNOQTASDS 112

Search completed: March 11, 2004, 19:13:10
 Job time : 171.3 secs

RESULT 2

```

; Sequence 2, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Escherichia coli
; US-09-741-873C-2

```

```

Query Match          54.7%; Score 414; DB 5; Length 131;
Best Local Similarity 69.9%; Pred. No. 1,1e-34;
Matches 79; Conservative 17; Mismatches 17; Indels 0; Gaps 0;

```

```

Qy 39 SGPSTLSIYOGSNNALALQSPARKSETTITSGYNGADVQAGDNSTIELTONGFR 98
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 19 SGPSEMLTYQYGGGNSALALQTDARNSDLTITOHGGNGADVQSGSDSIDLTQCGFG 78
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 99 NNATIDQWNAKSDITVQGYGANNALVNOTASDSSVWVRQYGVGNNAATNOY 151
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 79 NSATLIDQNGKNSMTYVQFGGNGAIVDQTAISNSVAVTVQVGFNNATIAHOY 131
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

RESULT 3

```

; Sequence 34, Application US/10004115B
; GENERAL INFORMATION:
; APPLICANT: ASAKO, HIROYUKI
; APPLICANT: MATSUMURA, KENJI
; APPLICANT: SHIMIZU, MASAYOSHI
; APPLICANT: ITO, NOBUYA
; APPLICANT: MAKITA, RYUHEI
; TITLE OF INVENTION: PROCESS FOR PRODUCING OPTICALLY ACTIVE
; FILE REFERENCE: 4-HALO-3-HYDROXYUTANOATE
; CURRENT APPLICATION NUMBER: US/10/004,115B
; CURRENT FILING DATE: 2001-12-06
; PRIOR APPLICATION NUMBER: JP 2000-372704
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: JP 2001-006144
; PRIOR FILING DATE: 2001-01-15
; PRIOR APPLICATION NUMBER: JP 2001-026594
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: JP 2001-175175
; PRIOR FILING DATE: 2001-06-11
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 34
; LENGTH: 348
; TYPE: PRT
; ORGANISM: Corynebacterium sp.

```

US-10-004-115B-34

```

Query Match          12.3%; Score 93; DB 6; Length 348;
Best Local Similarity 24.8%; Pred. No. 0.06;
Matches 40; Conservative 26; Mismatches 65; Indels 30; Gaps 6;

```

```

Qy 1 MKLITKAAFPATVYSGSALAGYDQVTRVVTHEMAHASPSTLSIYQY-GSANAALAL 59
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 191 LRHLSAATVIALDVNSADKL-----ELATKGAHEVV-LSDKDAEVRKRITSGGALVLT 244
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 60 QSDARSETTITSGYNGADV-----GGADNSTIELTONGFRNNATIDQWNAKNS--- 111
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 245 DFVGYQPTIDTMAVAGVSDVTIVIGIGQAHAKVGFQSPYEAASVTPYWGARELIE 304
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 112 -----DITVQYGGANNA-----LVNOTASDSSVWV 137
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 305 LIDLAHAGIFDISVETFSLDNGAEAVRILAACTLSGRAVIV 345
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

RESULT 4

```

; Sequence 40444, Application US/10767701
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 40444
; LENGTH: 358
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAV03-C6055_1.pep
; US-10-767-701-40444

```

```

Query Match          10.4%; Score 78.5; DB 6; Length 358;
Best Local Similarity 22.3%; Pred. No. 1.8;
Matches 35; Conservative 25; Mismatches 54; Indels 43; Gaps 6;

```

```

Qy 32 THEMHAAGPST---LSIYOGSNNALALQSPAR-----KSE 67
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 158 TDDRATGSLPBEITGNIPSVHTDNSQNDASENVQDATSATSDSSEHGTEAVHIETGLEDR 217
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 68 TTTTSGYNGADVQAGADNSTIELT-----QNGFRNNATID-----QWNAKNSDIT 114
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 218 SATASSGTGSGDDGKNSDSTPAEENETETASGDDEKMEGTCTTEAVHAKSEENSNSVS 277
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 115 VGOYV--GNNALVNOTASDSSVWVRQYGVGNNAATAN 149
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 278 TEAENSQDSSSGVNGSSEETS---NKGDGATVTSN 310
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

RESULT 5

```

; Sequence 44, Application PC/TUS0332645
; GENERAL INFORMATION:
; APPLICANT: Handfield, Martin
; APPLICANT: Hillman, Jeffrey
; APPLICANT: Progulske-Fox, Ann
; TITLE OF INVENTION: Identification of Actinobacillus actinomycetemcomitans Antigens
; FILE REFERENCE: MEH01-662B
; CURRENT APPLICATION NUMBER: PCT/US03/32645
; CURRENT FILING DATE: 2003-10-15
; NUMBER OF SEQ ID NOS: 234
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 44
; LENGTH: 1531

```



```

; TYPE: PRT
; ORGANISM: Actinobacillus actinomycetemcomitans
PCT-US03-32645-44

Query Match
Best Local Similarity 10.0%; Score 75.5; DB 1; Length 1511;
Matches 40; Conservative 27; Mismatches 57; Indels 43; Gaps 7;

QY 4 LKVAAPFAIVVSGSALAGVYDQVTRVTHEMAHAGSPDSTLSIYOGSANAALALQSDA 63
DB 293 LSKKADKALAVGTSALA---QKESAIARGYANAGIN-----ALISGANA 335

QY 64 RKSE-----TTTOSGYNGADVQGDANS-----TIELTONGFRNNATID 104
DB 336 KASQDNVVAIGKXATATSESG---SMAIGGAKSTFRNSALCTGTIVNSVDGQGSKFTAQ 392

QY 105 OMNAKNSDITVGYGNNALVNGTA---SDSSVMRYOYFGNNATA 148
DB 393 NYDANNGVAVAN-AGKERRIIVAGGRUDTDVANIQLKFTVNDILA 438

RESULT 6
US-10-767-701-42417
; Sequence 42417, Application US/10767701
; GENERAL INFORMATION:
; APPLICANT: Kovacic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 42417
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-C47268_1.pep
US-10-767-701-42417

Query Match
Best Local Similarity 9.8%; Score 74; DB 6; Length 386;
Matches 28; Conservative 25; Mismatches 55; Indels 20; Gaps 4;

QY 31 VTHEMAHAGSPDSTLSIYOGSANAALALQSDARKSETTTOSGYNGADVQ---GAD 86
DB 76 VPSESIHAATRG---GSDQAVSRHQPELKRITTTITGKHQTGSLGELLIPLD 128

QY 87 NSTIELTONGFRNNATIDOMNAKNSDITVGYGNNALVNGTASDSSVM-----VROV 140
DB 129 NENVNASS---NSSPTHTNRSVFTLTSSSSKQTYNQETSSDNNCKRPERNSTOGE 185

QY 141 GFGNNATA 148
DB 186 GISHSAAA 193

RESULT 7
US-10-093-037A-63
; Sequence 63, Application US/10093037A
; GENERAL INFORMATION:
; APPLICANT: Jay M. Short
; APPLICANT: Bylina, Edward
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Matchur, Eric J.
; APPLICANT: Lam, David E.
; TITLE OF INVENTION: ENZYMES HAVING GLUCOSIDASE ACTIVITY AND METHODS OF USE THEREOF
; FILE REFERENCE: 564462001402
; CURRENT APPLICATION NUMBER: US/10/093,037A
; CURRENT FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: US 09/910,579

```

```

; PRIOR FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: US 09/134,078
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 08/949,026
; PRIOR FILING DATE: 1997-10-10
; PRIOR APPLICATION NUMBER: US 60/056,916
; PRIOR FILING DATE: 1996-12-06
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 63
; LENGTH: 956
; TYPE: PRT
; ORGANISM: Bankia gouldi
US-10-093-037A-63

Query Match
Best Local Similarity 9.8%; Score 74; DB 6; Length 956;
Matches 37; Conservative 28; Mismatches 66; Indels 50; Gaps 8;

QY 11 AIVVSGSALAGVYD-----QVTRVYTH-EMAHAGSPDSTLSIYOGSANAALALQSD 62
DB 302 SVVQAGQVSGGLISNKLTFASGEIVKNIQNDTETSTGPKTT-----QCSTICIRAMET 357

QY 63 AKSETTTOSGYNGADVQGDNSTIEL--TONGFRNNATID----- 103
DB 358 AQAGDEIITAPGNVYFODKIQGAFNRSVYLYGSANGNSTPIILGESATNPVPSGLDY 417

QY 104 -----DOMNAKNSDITVGYG---GNNALVNGTASD---SSVMRYOYFGNN 145
DB 418 NNGYLSTIEGDYWNKIDIEFTKSGKIYDINSKGLKNLVYHDIGEEAIHLRD--GSSNN 476

QY 146 A 146
DB 477 S 477

RESULT 8
US-10-603-150-2
; Sequence 2, Application US/10603150
; GENERAL INFORMATION:
; APPLICANT: Robert G. Wisotzkey
; TITLE OF INVENTION: MAGE3-Like Gene Disruptions,
; FILE REFERENCE: R-1891
; CURRENT APPLICATION NUMBER: US/10/603,150
; CURRENT FILING DATE: 2003-06-24
; PRIOR APPLICATION NUMBER: US 60/391,205
; PRIOR FILING DATE: 2002-06-24
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1160
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-603-150-2

Query Match
Best Local Similarity 9.7%; Score 73.5; DB 6; Length 1160;
Matches 40; Conservative 24; Mismatches 55; Indels 45; Gaps 9;

QY 8 APAIVVSGSALAGVYDQVTRVTHEMAHAGSPDSTLSIYOGSANAALALQSDARKSE 67
DB 519 SPSGVLNSASFGALINT-----SAGFSTLN-----SSASFGSALSTSASFGG 562

QY 68 TTTOSGYNGADVQGDNSTIELTONG-----FRNNATIDOMNAKNSDITV--- 116
DB 563 VINGAGFGGALINT-----NATFGVINGSAGFGGAMNTNATFG--GALNSNAGFGGALS 615

QY 117 ---OYGG---NNAA-----LVNOTASDSSVMRYOYFGN--NATAN 149
DB 616 TSTNCGALNNSAGFGGAMNTSASFGVILNNSAGFGGALINTSAN 659

```

RESULT 9
US-10-767-701-47075
; Sequence 47075, Application US/10767701
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(5353)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 47075
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-C24509_1.psp
US-10-767-701-47075

Query Match 9.6%; Score 73; DB 6; Length 191;
Best Local Similarity 28.6%; Pred. No. 2.7;
Matches 36; Conservative 8; Mismatches 54; Indels 28; Gaps 3;

QY 7 AAFPAIVVSGSALAGVYDQVTRVVTTHMAHSGPDSITLTYOYGSANALALQSDARKS 66
DB 85 AAAAGAAGSAGAAAG-----SAGSAAAT-----GSGSAGAGSSTLLAS 123
QY 67 ETTTOSGNGADVCGADNSTELTONGFRNNATIDOMAKNSDITVCGYGGNNALV 126
DB 124 SAGAASSTGASASAGAGAAAS-----GTTAGAGVSTGASAVATTGVTSAVAIGD 176
QY 127 NOTASD 132
DB 177 DEKADD 182

RESULT 10
US-10-641-678-38
; Sequence 38, Application US/10641678
; GENERAL INFORMATION:
; APPLICANT: Day, Anthony, G.
; APPLICANT: Goedegebuurt, Frits
; APPLICANT: Gualfetti, Peter
; APPLICANT: Mitchinson, Colin
; APPLICANT: Neeffe, Paulien
; APPLICANT: Sandgren, Mats
; APPLICANT: Shaw, Andrew
; APPLICANT: Stahlberg, Jerry
; TITLE OF INVENTION: Novel Variant Hypocrea jecorina CBH1
; FILE REFERENCE: GC772-3
; CURRENT APPLICATION NUMBER: US/10/641,678
; CURRENT FILING DATE: 2003-08-15
; PRIOR APPLICATION NUMBER: US 60/458,853
; PRIOR FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: US 60/458,696
; PRIOR FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: US 60/456,368
; PRIOR FILING DATE: 2003-03-21
; PRIOR APPLICATION NUMBER: US 60/404,063
; PRIOR FILING DATE: 2002-08-16
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: FaastSeq for Windows Version 4.0
; SEQ ID NO 38
; LENGTH: 449
; TYPE: PRT
; ORGANISM: Phanerochaete chrysosporium
US-10-641-678-38

Query Match 9.5%; Score 72; DB 6; Length 449;
Best Local Similarity 26.1%; Pred. No. 11;

Matches 29; Conservative 14; Mismatches 46; Indels 22; Gaps 4;
QY 54 NAAL---ALQSDARKSETTTTOSGYNGADVCGADNSTELTONGFRNNATIDOMAKN 110
DB 157 NGALYFVAMDADGKSKR-----YFGNRAGAKYGYGCDSCPRDVGPIINGCAVYQGNMATS 212
QY 111 SDITVQYQY-----GNMALVNOTASDSSVMTROYGFGNNATN 149
DB 213 ATTGTGYSCTCELDIWEANSNAALTPHTCTNNA---QTRCSGNSCTSN 260

RESULT 11
PCT-US04-02188-185
; Sequence 185, Application PC/TUS0402188
; GENERAL INFORMATION:
; APPLICANT: Bayer Pharmaceuticals Corporation
; APPLICANT: Eweleigh, Deepa
; APPLICANT: Bigwood, Douglas
; APPLICANT: Taylor, Ian
; TITLE OF INVENTION: EXPRESSION PROFILES FOR COLON CANCER AND METHODS OF USE
; FILE REFERENCE: 5151
; CURRENT APPLICATION NUMBER: PCT/US04/02188
; CURRENT FILING DATE: 2004-01-23
; PRIOR APPLICATION NUMBER: 60/442,582
; PRIOR FILING DATE: 2003-01-24
; NUMBER OF SEQ ID NOS: 191
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 185
; LENGTH: 583
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US04-02188-185

Query Match 9.4%; Score 71.5; DB 1; Length 583;
Best Local Similarity 29.4%; Pred. No. 17;
Matches 32; Conservative 16; Mismatches 50; Indels 11; Gaps 5;

QY 27 VTRVVTTHMAHSGPDSITLTYOYGSANALALQSDARKSETTTTOSGYNGADVCGAD 86
DB 130 VARSTTYSLSHSEKRELLNVH---SANY-----SKVSFNEKAVTETSFNVSNNVGORGE 181
QY 87 NSTIELTON-GFRNNATIDOMAKNSDITV-GQYGGNN-NAALVNOTASD 132
DB 182 NSKSLTFNCSSTLNTITQSQHFLSPDSFVNNSHCANNELELVYCLSSD 230

RESULT 12
US-10-764-425-185
; Sequence 185, Application US/10764425
; GENERAL INFORMATION:
; APPLICANT: Bayer Pharmaceuticals Corporation
; APPLICANT: Eweleigh, Deepa
; APPLICANT: Bigwood, Douglas
; APPLICANT: Taylor, Ian
; TITLE OF INVENTION: EXPRESSION PROFILES FOR COLON CANCER AND METHODS OF USE
; FILE REFERENCE: 5151
; CURRENT APPLICATION NUMBER: US/10/764,425
; CURRENT FILING DATE: 2004-01-23
; PRIOR APPLICATION NUMBER: 60/442,582
; PRIOR FILING DATE: 2003-01-24
; NUMBER OF SEQ ID NOS: 191
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 185
; LENGTH: 583
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-764-425-185

Query Match 9.4%; Score 71.5; DB 6; Length 583;
Best Local Similarity 29.4%; Pred. No. 17;
Matches 32; Conservative 16; Mismatches 50; Indels 11; Gaps 5;
QY 27 VTRVVTTHMAHSGPDSITLTYOYGSANALALQSDARKSETTTTOSGYNGADVCGAD 86

Db 130 VRSTSYSLHASENELNVH---SANY-----SKVSFEKAVTSTSFNSVNVNGORGE 181

QY 87 NSTIETLON-GFRNNATIDOMNAKNSDITY-GQYGGNN-AALVNOTASD 132

Db 182 NSTSLTPKCSSTINITOSQIHFLSPDSFVNSHGANNELELVTCISSD 230

RESULT 13

US-10-771-241-364

; Sequence 364, Application US/10771241

; GENERAL INFORMATION:

; APPLICANT: Zyklich, Judith

; APPLICANT: Forsyth, R. Allyn

; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN

; FILE REFERENCE: ELITRA 001C1

; CURRENT APPLICATION NUMBER: US/10/771,241

; CURRENT FILING DATE: 2004-02-03

; PRIOR APPLICATION NUMBER: 09/492,709

; PRIOR FILING DATE: 2000-01-27

; PRIOR APPLICATION NUMBER: 60/117,405

; PRIOR FILING DATE: 1999-01-27

; NUMBER OF SEQ ID NOS: 485

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 364

; LENGTH: 878

; TYPE: PRT

; ORGANISM: E. Coli

US-10-771-241-364

Query Match

Best local Similarity 9.4%; Score 71.5; DB 6; Length 878;

Matches 34; Conservative 19; Mismatches 68; Indels 29; Gaps 5;

QY 24 DQLVTRVVTHEMAHSGPDSSTLSIYQGSANALALQSDARKSETTI----- 70

Db 591 DQMLANLVNIPFSHMRSDSK--SQMRHASASYSMSHDINGMTVLGAGYGTILEDNNL 647

QY 71 ----TOSGNGAGVAGGADNSTIELTONGFRNNATIDOMNAKNSDITVGOYGN----- 121

Db 648 SYSVQGVYAGGGD-GNSGSTGYTLNLYRGQYGNANIGY--SHSDIKQLYGVSGVLAH 704

QY 122 -NAALVNOTASDSSVWVRQYGFNNATYANQ 150

Db 705 ANGVTIGQPLNDITVVLVKAPGADAYENG 734

RESULT 14

US-10-767-701-44555

; Sequence 44555, Application US/10767701

; GENERAL INFORMATION:

; APPLICANT: Kovacic, David K.

; APPLICANT: Zhou, Yihua

; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

; FILE REFERENCE: 38-21(5353)B

; CURRENT APPLICATION NUMBER: US/10/767,701

; CURRENT FILING DATE: 2004-01-29

; NUMBER OF SEQ ID NOS: 63128

; SEQ ID NO 44555

; LENGTH: 281

; TYPE: PRT

; ORGANISM: Sorghum bicolor

; FEATURE:

; OTHER INFORMATION: Clone ID: SORBI-28MAY03-C21806_1.dep

US-10-767-701-44555

Query Match

Best local Similarity 9.3%; Score 70.5; DB 6; Length 281;

Matches 23; Conservative 13; Mismatches 31; Indels 15; Gaps 2;

QY 9 PAIVVSGSALAGVTDQLVTRVVTHEMAHSGPDSSTLSIY-----QYGSAN 54

Db 10 FLVLVPGPSLTGVFSDSPFSLVSGFHLFVSSPVSRVLFCTEAYFLRLCTVLVRYFSAV 69

QY 55 A-ALALQSDARKSETTIQSGY 75

Db 70 AFRLCQVTCHQTKTNMLQPGF 91

RESULT 15

US-10-451-467A-462

; Sequence 462, Application US/10451467A

; GENERAL INFORMATION:

; APPLICANT: CONTRERAS, ROLAND HENRI

; APPLICANT: EBERHARDT, INES

; APPLICANT: LUYTEN, WALTER HERMAN MARIA LOUIS

; APPLICANT: BEKKMAN, RIEKA JOSEPHINA

; TITLE OF INVENTION: BAX-RESPONSIVE GENES FOR DRUG TARGET IDENTIFICATION IN

; FILE REFERENCE: JAB-1667

; CURRENT APPLICATION NUMBER: US/10/451,467A

; CURRENT FILING DATE: 2003-06-19

; PRIOR APPLICATION NUMBER: EP 00870318.3

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: EP 01870002.1

; PRIOR FILING DATE: 2001-01-04

; PRIOR APPLICATION NUMBER: EP 01870003.9

; PRIOR FILING DATE: 2001-01-09

; NUMBER OF SEQ ID NOS: 732

; SOFTWARE: Patentin version 3.1

; SEQ ID NO 462

; LENGTH: 829

; TYPE: PRT

; ORGANISM: Candida albicans

US-10-451-467A-462

Query Match

Best local Similarity 9.2%; Score 70; DB 6; Length 829;

Matches 32; Conservative 17; Mismatches 44; Indels 24; Gaps 6;

QY 39 SGPDSSTLS---IYQGSANALALQSDARKSETTIQSGYNGAGVAGGADNSTIELTON 95

Db 635 NGPGSVTSTPLILNCGSANMKLS-----NLSTISVTGTGAGTVTGTG3-----APD 660

QY 96 GFRNNATIDOMNAKNSDITVGOYGGNNALVNOTASDS-SVWVRQYGFNNATYANQ 151

Db 681 GFPPIGSPDETTPNS-----IFGHTRGLPSSRLSESQSVLYQYVAIAD-D-TVGDY 731

Search completed: March 11, 2004, 19:14:16

Job time : 6.3 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:33:51 / Search time 171.3 Seconds

(without alignments)
860.386 Million cell updates/sec

Title: US-09-543-407-18

Perfect score: 780
Sequence: 1 MKLLRVAAFAALVSGSALA.....DSSVNVROYFGNNATANY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 6019581 seqs, 976053577 residues

Total number of hits satisfying chosen parameters: 6019581

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Pending Patents_AA_Main:*

1: /cgn2_6/ptodata/2/paa/US006_COMB.pep.*
2: /cgn2_6/ptodata/2/paa/US006_COMB.pep.*
3: /cgn2_6/ptodata/2/paa/US007_COMB.pep.*
4: /cgn2_6/ptodata/2/paa/US008_COMB.pep.*
5: /cgn2_6/ptodata/2/paa/US008_COMB.pep.*
6: /cgn2_6/ptodata/2/paa/US008_COMB.pep.*
7: /cgn2_6/ptodata/2/paa/US008_COMB.pep.*
8: /cgn2_6/ptodata/2/paa/US008_COMB.pep.*
9: /cgn2_6/ptodata/2/paa/US008_COMB.pep.*
10: /cgn2_6/ptodata/2/paa/US008_COMB.pep.*
11: /cgn2_6/ptodata/2/paa/US008_COMB.pep.*
12: /cgn2_6/ptodata/2/paa/US008_COMB.pep.*
13: /cgn2_6/ptodata/2/paa/US009_COMB.pep.*
14: /cgn2_6/ptodata/2/paa/US009_COMB.pep.*
15: /cgn2_6/ptodata/2/paa/US009_COMB.pep.*
16: /cgn2_6/ptodata/2/paa/US009_COMB.pep.*
17: /cgn2_6/ptodata/2/paa/US009_COMB.pep.*
18: /cgn2_6/ptodata/2/paa/US009_COMB.pep.*
19: /cgn2_6/ptodata/2/paa/US009_COMB.pep.*
20: /cgn2_6/ptodata/2/paa/US009_COMB.pep.*
21: /cgn2_6/ptodata/2/paa/US009_COMB.pep.*
22: /cgn2_6/ptodata/2/paa/US009_COMB.pep.*
23: /cgn2_6/ptodata/2/paa/US009_COMB.pep.*
24: /cgn2_6/ptodata/2/paa/US009_COMB.pep.*
25: /cgn2_6/ptodata/2/paa/US009_COMB.pep.*
26: /cgn2_6/ptodata/2/paa/US009_COMB.pep.*
27: /cgn2_6/ptodata/2/paa/US010_COMB.pep.*
28: /cgn2_6/ptodata/2/paa/US010_COMB.pep.*
29: /cgn2_6/ptodata/2/paa/US010_COMB.pep.*
30: /cgn2_6/ptodata/2/paa/US010_COMB.pep.*
31: /cgn2_6/ptodata/2/paa/US010_COMB.pep.*
32: /cgn2_6/ptodata/2/paa/US010_COMB.pep.*
33: /cgn2_6/ptodata/2/paa/US010_COMB.pep.*

Prod. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
------------	-------	-------------	--------	----	-------------

1	780	100.0	151	19	US-09-543-407-18	Sequence 18, Appl
2	691	88.6	151	19	US-09-543-407-5	Sequence 5, Appl
3	686	87.9	151	6	US-08-233-642A-57	Sequence 57, Appl
4	675	86.5	151	19	US-09-543-407-26	Sequence 26, Appl
5	622	79.7	151	19	US-09-543-407-20	Sequence 20, Appl
6	618	79.2	151	19	US-09-543-407-16	Sequence 16, Appl
7	613	78.6	151	19	US-09-543-407-12	Sequence 12, Appl
8	611	78.3	151	19	US-09-543-407-14	Sequence 14, Appl
9	608	77.9	151	19	US-09-543-407-24	Sequence 24, Appl
10	604	77.4	131	19	US-09-543-407-11	Sequence 31, Appl
11	601	77.1	151	19	US-09-543-407-28	Sequence 28, Appl
12	600	76.9	151	19	US-09-543-407-22	Sequence 22, Appl
13	599	76.8	151	19	US-09-543-407-30	Sequence 30, Appl
14	523	67.1	151	19	US-09-543-407-7	Sequence 7, Appl
15	520	66.7	151	13	US-08-978-878-4	Sequence 4, Appl
16	520	66.7	151	21	US-09-741-873B-4	Sequence 4, Appl
17	518	66.4	151	33	US-60-352-946-2	Sequence 2, Appl
18	518	66.4	151	33	US-60-444-371-2	Sequence 2, Appl
19	506	64.9	120	6	US-08-233-642A-55	Sequence 55, Appl
20	483	61.9	109	19	US-09-543-407-24	Sequence 34, Appl
21	457	58.6	158	16	US-09-252-691C-5834	Sequence 5834, Ap
22	457	58.6	158	16	US-09-252-691C-5834	Sequence 5834, Ap
23	457	58.6	158	30	US-10-417-886-5834	Sequence 5834, Ap
24	444	56.9	131	13	US-08-978-878-2	Sequence 2, Appl
25	444	56.9	131	21	US-09-741-873B-2	Sequence 2, Appl
26	359.5	46.1	109	19	US-09-543-407-35	Sequence 35, Appl
27	276	35.4	68	19	US-09-543-407-37	Sequence 37, Appl
28	250.5	32.1	70	19	US-09-543-407-32	Sequence 32, Appl
29	175	22.4	48	19	US-09-543-407-39	Sequence 39, Appl
30	116	14.9	186	16	US-09-252-691C-5833	Sequence 5833, Ap
31	116	14.9	186	16	US-09-252-691C-5833	Sequence 5833, Ap
32	116	14.9	186	30	US-10-417-886-5833	Sequence 5833, Ap
33	108.5	13.9	445	29	US-10-369-493-20638	Sequence 20638, A
34	108.5	13.9	445	33	US-60-360-039-20638	Sequence 20638, A
35	107.5	13.8	151	19	US-09-543-407-8	Sequence 8, Appl
36	106.5	13.7	151	19	US-09-543-407-6	Sequence 6, Appl
37	101	12.9	257	33	US-60-173-464-21553	Sequence 21553, A
38	101	12.9	262	20	US-09-614-150A-25818	Sequence 25818, A
39	101	12.9	262	20	US-09-614-150A-25818	Sequence 25818, A
40	101	12.9	262	33	US-60-191-637-25957	Sequence 25957, A
41	101	12.9	262	33	US-60-191-637-25957	Sequence 25957, A
42	100	12.8	1308	27	US-10-179-131-5148	Sequence 5148, Ap
43	99.5	12.8	1249	30	US-10-455-719-358	Sequence 357, App
44	99.5	12.8	1249	33	US-60-385-568-357	Sequence 357, App
45	99.5	12.8	1249	33	US-60-446-775-358	Sequence 358, App

ALIGNMENTS

RESULT 1
US-09-543-407-18
Sequence 18, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543,407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastrSeq for Windows Version 4.0
SEQ ID NO 18
LENGTH: 151
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
Sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.

US-09-543-407-18

Query Match 100.0%; Score 780; DB 19; Length 151;
Best Local Similarity 100.0%; Pred. No. 3.7e-75;
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLKVAAFAAIVVSGSALAGVVPWGGGNNHNGGNSGPDYDOLVTRVVTHEMAHALQ 60
DB 1 MKLKVAAFAAIVVSGSALAGVVPWGGGNNHNGGNSGPDYDOLVTRVVTHEMAHALQ 60
QY 61 SPARKSETTTTOSGYGNADVGQADNSTIELTONGFRNNATIDOWNAKNSDITVGQYGG 120
DB 61 SPARKSETTTTOSGYGNADVGQADNSTIELTONGFRNNATIDOWNAKNSDITVGQYGG 120
QY 121 NNAALVNOTASDSSVMVROVGFNNATANQY 151
DB 121 NNAALVNOTASDSSVMVROVGFNNATANQY 151

RESULT 2

US-09-543-407-5
Sequence 5, Application US/09543407

GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIBRILLAR SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5
LENGTH: 151
TYPE: PRT
ORGANISM: Salmonella enteritidis
US-09-543-407-5

Query Match 88.6%; Score 691; DB 19; Length 151;
Best Local Similarity 90.7%; Pred. No. 1.4e-65;
Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 1 MKLKVAAFAAIVVSGSALAGVVPWGGGNNHNGGNSGPDYDOLVTRVVTHEMAHALQ 60
DB 1 MKLKVAAFAAIVVSGSALAGVVPWGGGNNHNGGNSGPDYDOLVTRVVTHEMAHALQ 60
QY 61 SPARKSETTTTOSGYGNADVGQADNSTIELTONGFRNNATIDOWNAKNSDITVGQYGG 120
DB 61 SPARKSETTTTOSGYGNADVGQADNSTIELTONGFRNNATIDOWNAKNSDITVGQYGG 120
QY 121 NNAALVNOTASDSSVMVROVGFNNATANQY 151
DB 121 NNAALVNOTASDSSVMVROVGFNNATANQY 151

RESULT 3

US-08-233-642A-57
Sequence 57, Application US/08233642A

GENERAL INFORMATION:
APPLICANT: Kay, William W.
APPLICANT: Collinson, S. Karen
APPLICANT: Clouthier, Sharon C.
APPLICANT: Doran, James L.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
TITLE OF INVENTION: BASED VACCINES
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed and Berry
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle

STATE: Washington

COUNTRY: U.S.A.

ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/233,642A

FILING DATE: 26-APR-1994

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: King, Joshua

REGISTRATION NUMBER: 35,570

REFERENCE/DOCKET NUMBER: 920043.403C3

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

TELEX: 3723836 SEEDANBERY

INFORMATION FOR SEQ ID NO: 57:

SEQUENCE CHARACTERISTICS:

LENGTH: 151 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-233-642A-57

Query Match 87.9%; Score 686; DB 6; Length 151;
Best Local Similarity 90.1%; Pred. No. 4.8e-65;
Matches 136; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

QY 1 MKLKVAAFAAIVVSGSALAGVVPWGGGNNHNGGNSGPDYDOLVTRVVTHEMAHALQ 60
DB 1 MKLKVAAFAAIVVSGSALAGVVPWGGGNNHNGGNSGPDYDOLVTRVVTHEMAHALQ 60
QY 61 SPARKSETTTTOSGYGNADVGQADNSTIELTONGFRNNATIDOWNAKNSDITVGQYGG 120
DB 61 SPARKSETTTTOSGYGNADVGQADNSTIELTONGFRNNATIDOWNAKNSDITVGQYGG 120
QY 121 NNAALVNOTASDSSVMVROVGFNNATANQY 151
DB 121 NNAALVNOTASDSSVMVROVGFNNATANQY 151

RESULT 4

US-09-543-407-26
Sequence 26, Application US/09543407

GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIBRILLAR SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 26
LENGTH: 151
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afaA
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-26

Query Match 86.5%; Score 675; DB 19; Length 151;
Best Local Similarity 81.9%; Pred. No. 7.3e-64;
Matches 136; Conservative 0; Mismatches 0; Indels 30; Gaps 2;

```

QY      1 MKLKVAAFAIIVSGSALAGVVPQWGGGNNHNGGNSGPD-----YDQ 45
      1 MKLKVAAFAIIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLIYQGSANALAYQ 60
DB      46 LVTRVVTHEMAHALQSDARKSETTITQSGYNGADVQGGADNSTIELTONGFRNNATIQ 105
      61 LVTRVVTHEMAHA-----GYGNADVQGGADNSTIELTONGFRNNATIQ 105
QY      106 WNAKNSDITVQGYGNNALVNOTASDSVWVRQVFGNNATANY 151
      106 WNAKNSDITVQGYGNNALVNOTASDSVWVRQVFGNNATANY 151
DB

```

```

RESULT 5
US-09-543-407-20
; Sequence 20, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GPe3 of Leishmania major.
US-09-543-407-20

```

```

Query Match      79.7%; Score 622; DB 19; Length 151;
Best local Similarity 73.6%; Pred. No. 3.6e-58;
Matches 128; Conservative 0; Mismatches 0; Indels 46; Gaps 2;

QY      1 MKLKVAAFAIIVSGSALAGVVPQWGGGNNHNGGNSGPD----- 42
      1 MKLKVAAFAIIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLIYQGSANALAYQ 60
DB      43 -----YDQLVTRVVTHEMAHALQSDARKSETTITQSGYNGADVQGGADNSTIELTONGF 97
      61 SPARKYDQLVTRVVTHEMAHA-----GQGANSTIELTONGF 97
QY      98 RNNATIDQNAKNSDITVQGYGNNALVNOTASDSVWVRQVFGNNATANY 151
      98 RNNATIDQNAKNSDITVQGYGNNALVNOTASDSVWVRQVFGNNATANY 151
DB

```

```

RESULT 6
US-09-543-407-16
; Sequence 16, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 151

```

```

; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GPe3 of Leishmania major.
US-09-543-407-16

```

```

Query Match      79.2%; Score 618; DB 19; Length 151;
Best local Similarity 76.6%; Pred. No. 9.8e-58;
Matches 131; Conservative 0; Mismatches 0; Indels 40; Gaps 2;

```

```

QY      1 MKLKVAAFAIIVSGSALAGVVPQWGGGNNHNGGNSGPDYDQLVTRVVTHEMAH--- 57
      1 MKLKVAAFAIIVSGSALAGV-----YDQLVTRVVTHEMAHAG 40
DB      58 -----ALQSDARKSETTITQSGYNGADVQGGADNSTIELTONGFRNN 100
      41 PDSTLIYQGSANALALQSDARKSETTITQSGYNGADVQGGADNSTIELTONGFRNN 100
QY      101 ATIDQNAKNSDITVQGYGNNALVNOTASDSVWVRQVFGNNATANY 151
      101 ATIDQNAKNSDITVQGYGNNALVNOTASDSVWVRQVFGNNATANY 151
DB

```

```

RESULT 7
US-09-543-407-12
; Sequence 12, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GPe3 of Leishmania major.
US-09-543-407-12

```

```

Query Match      78.6%; Score 613; DB 19; Length 151;
Best local Similarity 80.8%; Pred. No. 3.4e-57;
Matches 122; Conservative 7; Mismatches 22; Indels 0; Gaps 0;

QY      1 MKLKVAAFAIIVSGSALAGVVPQWGGGNNHNGGNSGPDYDQLVTRVVTHEMAHALQ 60
      1 MKLKVAAFAIIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLIYQGSANALAYQ 60
DB      61 SPARKSETTITQSGYNGADVQGGADNSTIELTONGFRNNATIDQNAKNSDITVQGYG 120
      61 SPARKSETTITQSGYNGADVQGGADNSTIELTONGFRNNATIDQNAKNSDITVQGYG 120
QY      121 NNAALVNOTASDSVWVRQVFGNNATANY 151
      121 NNAALVNOTASDSVWVRQVFGNNATANY 151
DB

```

```

RESULT 8
US-09-543-407-14
; Sequence 14, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.

```

APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543,407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 14
LENGTH: 151
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-14

Query Match 78.3%; Score 611; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 5.5e-57;
Matches 123; Conservative 6; Mismatches 22; Indels 0; Gaps 0;

QY 1 MKLKVAFPAIVVSGSALAGVVPQWGGGNNHNGGNSGPDYDQVTVRVVTHEMAHALQ 60
DB 1 MKLKVAFPAIVVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
QY 61 SDARKSETTITGSGYGADVGAGADNSTIELTONGFRNNATIDQNNAKNSDITVGQYGG 120
DB 61 SDARKSETTITGSGYGADVGAGADNSTIELTONGFRNNATIDQNNAKNSDITVGQYGG 120
QY 121 NNAALVNOTASDSSVMRVQVGFNNATANY 151
DB 121 LVTRVTHEMAHASVMRVQVGFNNATANY 151

RESULT 9
US-09-543-407-24
Sequence 24, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543,407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 24
LENGTH: 151
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-24

Query Match 77.9%; Score 608; DB 19; Length 151;
Best Local Similarity 82.1%; Pred. No. 1.2e-56;
Matches 124; Conservative 5; Mismatches 22; Indels 0; Gaps 0;
QY 1 MKLKVAFPAIVVSGSALAGVVPQWGGGNNHNGGNSGPDYDQVTVRVVTHEMAHALQ 60
DB 1 MKLKVAFPAIVVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
QY 61 SDARKSETTITGSGYGADVGAGADNSTIELTONGFRNNATIDQNNAKNSDITVGQYGG 120
DB 61 SDARKSETTITGSGYGADVGAGADNSTIELTONGFRNNATIDQNNAKNSDITVGQYGG 120

QY 121 NNAALVNOTASDSSVMRVQVGFNNATANY 151
DB 121 HEMAHANQTSASSVMRVQVGFNNATANY 151

RESULT 10
US-09-543-407-31
Sequence 31, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543,407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 31
LENGTH: 131
TYPE: PRT
ORGANISM: Salmonella enteritidis
US-09-543-407-31

Query Match 77.4%; Score 604; DB 19; Length 131;
Best Local Similarity 89.3%; Pred. No. 2.6e-56;
Matches 117; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 21 GVVPQWGGGNNHNGGNSGPDYDQVTVRVVTHEMAHALQSDARKSETTITGSGYGAD 80
DB 1 GVVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQSDARKSETTITGSGYGAD 60
QY 81 VQAGADNSTIELTONGFRNNATIDQNNAKNSDITVGQYGGNNAALVNOTASDSSVMRVQ 140
DB 61 VQAGADNSTIELTONGFRNNATIDQNNAKNSDITVGQYGGNNAALVNOTASDSSVMRVQ 120
QY 141 GFNNATANY 151
DB 121 GFNNATANY 131

RESULT 11
US-09-543-407-28
Sequence 28, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543,407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 28
LENGTH: 151
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-28

Query Match 77.1%; Score 601; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 6.6e-56;
Matches 122; Conservative 4; Mismatches 25; Indels 0; Gaps 0;
QY 1 MKLKVAFPAIVVSGSALAGVVPQWGGGNNHNGGNSGPDYDQVTVRVVTHEMAHALQ 60


```
Db 1 MLLKVAAPAAIIVSGSALAGVPPQWGGGNNHGGNSGPDSTLSIYOGSANAALALQ 60
QY 61 SPARKSETTITGSGYNGADVGGADNSTIELTONGFRNNATIDQWNAKSDITTVGOYGG 120
Db 61 SPARKSETTITGSGYNGADVGGADNSTIELTONGFRNNATIDQWNAKSDITTVGOYGG 120
QY 121 NNAALVNQTASDSSVWVRQVGFNNATANQY 151
Db 121 NNAALVNQTASDSSVWVRQVGFNNATANQY 151

RESULT 12
US-09-543-407-22
; Sequence 22, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FaastSeq for Windows Version 4.0
; SEQ ID NO 22
; LENGTH: 151
; TYPE: PRF
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
US-09-543-407-22

Query Match 76.9%; Score 600; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 8.4e-56;
Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;

QY 1 MLLKVAAPAAIIVSGSALAGVPPQWGGGNNHGGNSGPDSTLSIYOGSANAALALQ 60
Db 1 MLLKVAAPAAIIVSGSALAGVPPQWGGGNNHGGNSGPDSTLSIYOGSANAALALQ 60
QY 61 SPARKSETTITGSGYNGADVGGADNSTIELTONGFRNNATIDQWNAKSDITTVGOYGG 120
Db 61 SPARKSETTITGSGYNGADVGGADNSTIELTONGFRNNATIDQWNAKSDITTVGOYGG 120
QY 121 NNAALVNQTASDSSVWVRQVGFNNATANQY 151
Db 121 NNAALVNQTASDSSVWVRQVGFNNATANQY 151

RESULT 13
US-09-543-407-30
; Sequence 30, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FaastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 151
; TYPE: PRF
; ORGANISM: Artificial Sequence
```

```
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GPe3 of Leishmania major.
US-09-543-407-30

Query Match 76.8%; Score 599; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 1.1e-55;
Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;

QY 1 MLLKVAAPAAIIVSGSALAGVPPQWGGGNNHGGNSGPDSTLSIYOGSANAALALQ 60
Db 1 MLLKVAAPAAIIVSGSALAGVPPQWGGGNNHGGNSGPDSTLSIYOGSANAALALQ 60
QY 61 SPARKSETTITGSGYNGADVGGADNSTIELTONGFRNNATIDQWNAKSDITTVGOYGG 120
Db 61 SPARKSETTITGSGYNGADVGGADNSTIELTONGFRNNATIDQWNAKSDITTVGOYGG 120
QY 121 NNAALVNQTASDSSVWVRQVGFNNATANQY 151
Db 121 NNAALVNQTASDSSVWVRQVGFNNATANQY 151

RESULT 14
US-09-543-407-7
; Sequence 7, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FaastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 151
; TYPE: PRF
; ORGANISM: Escherichia coli
US-09-543-407-7

Query Match 67.1%; Score 523; DB 19; Length 151;
Best Local Similarity 68.9%; Pred. No. 1.6e-47;
Matches 104; Conservative 20; Mismatches 27; Indels 0; Gaps 0;

QY 1 MLLKVAAPAAIIVSGSALAGVPPQWGGGNNHGGNSGPDSTLSIYOGSANAALALQ 60
Db 1 MLLKVAAPAAIIVSGSALAGVPPQWGGGNNHGGNSGPDSTLSIYOGSANAALALQ 60
QY 61 SPARKSETTITGSGYNGADVGGADNSTIELTONGFRNNATIDQWNAKSDITTVGOYGG 120
Db 61 TDARNSDLITTYHGGNGADVGGSDSSIDLTRGFSGSATLTDQWNGKNSERTVTFQGG 120
QY 121 NNAALVNQTASDSSVWVRQVGFNNATANQY 151
Db 121 GNGAAMDQTASNSVWVRQVGFNNATANQY 151

RESULT 15
US-08-978-878-4
; Sequence 4, Application US/08978878
; GENERAL INFORMATION:
; APPLICANT: NORMARK, Staffan
; APPLICANT: OLSEN, Arne
; TITLE OF INVENTION: FIBROECTIN BINDING PROTEIN AS WELL AS ITS PREPARATION
; FILE REFERENCE: 012889-081
; CURRENT APPLICATION NUMBER: US/08/978,878
; CURRENT FILING DATE: 1997-11-26
; EARLIER APPLICATION NUMBER: SE 8801723-1
; EARLIER FILING DATE: 1988-05-06
```

```

; EARLIER APPLICATION NUMBER: US 07/347,189
; EARLIER FILING DATE: 1989-05-04
; EARLIER APPLICATION NUMBER: US 07/789,437
; EARLIER FILING DATE: 1991-11-06
; EARLIER APPLICATION NUMBER: US 07/970,846
; EARLIER FILING DATE: 1992-11-03
; EARLIER APPLICATION NUMBER: US 08/187,865
; EARLIER FILING DATE: 1994-01-28
; EARLIER APPLICATION NUMBER: US 08/318,519
; EARLIER FILING DATE: 1994-10-05
; EARLIER APPLICATION NUMBER: US 08/495,959
; EARLIER FILING DATE: 1995-06-28
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-08-978-878-4

```

```

Query Match      66.7%; Score 520; DB 13; Length 151;
Best Local Similarity 68.2%; Pred. No. 3.4e-47;
Matches 103; Conservative 21; Mismatches 27; Indels 0; Gaps 0;

```

```

QY      1 MKLIKVAFAAIVVSGSALAGVFPQWGGGNNHGGNSGPDYDQLVTRVVTHEMAHALQ 60
Db      1 MKLIKVAFAAIVVSGSALAGVFPQYGGGNNHGGGNNSGPNSSELINITYOGGNSALALQ 60
QY      61 SDARKSETTTTQSGYGAGADVGAGADNSTIELTONGFRNNATIDQNNAKNSDITVGQYGG 120
Db      61 TDARNSDLTTTQHGAGAGADVGAGSDSIDLTORFGNSATLIDQNNKNSSEMTVKQFGG 120
QY      121 NNALVNOTASDSVMTROYGFGNNATATANOY 151
Db      121 GNGAAVDQTAASSSVNTQVGFNNATATHQY 151

```

Search completed: March 11, 2004, 19:13:11
 Job time : 172.3 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:35:05 ; Search time 5.3 Seconds
(without alignments)
376.014 Million cell updates/sec

Title: US-09-543-407-18

Perfect score: 780
Sequence: 1 MTLKVAFAIAVSGSALA.....DSVMVRYQFGNNATANY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 125546 seqs, 13197846 residues

Total number of hits satisfying chosen parameters: 125546

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents_AA_New: *
1: /cgn2_6/prodata/1/paa/PCT_NEW_COMB.pep: *
2: /cgn2_6/prodata/1/paa/US06_NEW_COMB.pep: *
3: /cgn2_6/prodata/1/paa/US07_NEW_COMB.pep: *
4: /cgn2_6/prodata/1/paa/US08_NEW_COMB.pep: *
5: /cgn2_6/prodata/1/paa/US09_NEW_COMB.pep: *
6: /cgn2_6/prodata/1/paa/US10_NEW_COMB.pep: *
7: /cgn2_6/prodata/1/paa/US60_NEW_COMB.pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	520	66.7	151	US-09-741-873C-4	Sequence 4, Appl
2	444	56.9	131	US-09-741-873C-2	Sequence 2, Appl
3	82.5	10.6	434	US-10-045-674A-594	Sequence 594, App
4	81.5	10.4	424	US-10-045-674A-591	Sequence 527, App
5	81.5	10.4	533	US-10-045-674A-527	Sequence 527, App
6	79.5	10.2	956	US-10-093-037A-63	Sequence 63, Appl
7	78.5	10.1	480	PCT-US04-05654-1498	Sequence 1498, App
8	77.5	9.9	234	US-10-767-701-45603	Sequence 45603, A
9	77.5	9.9	443	US-10-100-683-7608	Sequence 7608, App
10	76.5	9.8	179	US-10-767-701-35342	Sequence 35342, A
11	76.5	9.8	348	US-10-004-115B-34	Sequence 34, Appl
12	74.5	9.6	430	US-10-451-467A-314	Sequence 314, Appl
13	74	9.5	305	PCT-US04-05654-1406	Sequence 1406, App
14	73	9.4	310	PCT-US04-05654-1412	Sequence 1412, App
15	73	9.4	386	US-10-767-701-42417	Sequence 42417, A
16	73	9.4	1327	PCT-US04-02338-49	Sequence 49, Appl
17	72	9.2	654	PCT-US04-02188-156	Sequence 156, App
18	72	9.2	654	US-10-764-425-156	Sequence 156, App
19	71.5	9.2	386	US-10-100-683-10326	Sequence 10326, A
20	71.5	9.2	386	US-10-100-683-10327	Sequence 10327, A
21	71.5	9.2	772	US-10-417-884A-5724	Sequence 5724, App
22	71	9.1	391	PCT-US04-05654-336	Sequence 336, App
23	70.5	9.0	173	US-10-767-701-39649	Sequence 39649, A
24	70	9.0	1160	US-10-603-150-2	Sequence 2, Appl
25	69.5	8.9	449	US-10-641-678-38	Sequence 38, Appl
26	69	8.8	245	US-10-779-461-19	Sequence 19, Appl

27	69	8.8	574	1	PCT-US04-05654-334	Sequence 334, App
28	69	8.8	633	6	US-10-781-979-11	Sequence 11, Appl
29	69	8.8	633	6	US-10-782-020-8	Sequence 8, Appl
30	69	8.8	1135	1	PCT-US04-05092-23	Sequence 23, Appl
31	68.5	8.8	346	1	PCT-US04-0242-58	Sequence 58, Appl
32	68	8.7	364	1	PCT-US04-05654-1756	Sequence 1756, App
33	67.5	8.7	292	1	PCT-US04-05654-2080	Sequence 2080, App
34	67.5	8.7	358	6	US-10-767-701-40444	Sequence 40444, A
35	67.5	8.7	433	1	PCT-US04-05654-2586	Sequence 2586, App
36	67	8.6	310	6	US-10-786-850-4	Sequence 4, Appl
37	67	8.6	321	1	PCT-US04-05654-590	Sequence 590, App
38	67	8.6	878	6	US-10-771-241-364	Sequence 364, App
39	66.5	8.5	147	6	US-10-767-701-38351	Sequence 58351, A
40	66.5	8.5	317	1	PCT-US04-05654-2038	Sequence 2038, App
41	66.5	8.5	317	1	PCT-US04-05654-2370	Sequence 2370, App
42	66.5	8.5	342	6	US-10-767-701-46086	Sequence 46086, A
43	66	8.5	221	6	US-10-767-701-41435	Sequence 41435, A
44	66	8.5	245	6	US-10-779-461-59	Sequence 59, Appl
45	66	8.5	442	1	PCT-US04-05654-338	Sequence 338, App

ALIGNMENTS

RESULT 1
US-09-741-873C-4
; Sequence 4, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibrinectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873C-4

Query Match 66.7%; Score 520; DB 5; Length 151;
Best local similarity 68.2%; Pred. No. 4.8e-40;
Matches 103; Conservative 21; Mismatches 27; Indels 0; Gaps 0;

QY	1	MTLKVAFAIAVSGSALAGVVPQWGGGNNNGGNSGPDYDQVTRVVTHEMAHAIQ	60
DB	1	MTLKVAFAIAVSGSAVAVPVQGGGNGGGGNSGPNSELNIVYGGGNSALAIQ	60
QY	61	SPARSETTITGSGGNGADVGQGDNSITETLTONCFRNATIDQNAKNSDITVOYGQ	120
DB	61	TPARRSDLTITGCGGNGADVGQSDSSIDILTRKFGNSATVDQNGKNSKTVKQFG	120
QY	121	NNAAVNTASDSSVMVRYQFGNNATANY	151
DB	121	GNGAVDQFASNSVMVRYQFGNNATANY	151

RESULT 2

```

; US-09-741-873C-2
; Sequence 2, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Escherichia coli
; US-09-741-873C-2

```

Query Match

```

Best Local Similarity 56.9%; Score 444; DB 5; Length 131;
Best Local Similarity 65.6%; Pred. No. 2.7e-33;
Matches 86; Conservative 19; Mismatches 26; Indels 0; Gaps 0;

```

```

QY 21 GVPPWGGGNNHNGGSSGPDYQVLTVRVYTHMAHALQSDARKSETTTTQSGYGNCA 80
DB 1 GVPPYGGGNNHNGGSSGPNSEINTYQYGGNSALALQTDANSDLITTHGGGNGAD 60
QY 81 VGGADNSTIELTNGFRNNATIDQWNAKNSDITVGYGNNALVNOTASDSSVWVRQ 140
DB 61 VGGSDSSIDLTRGFNSATLQWNGKNSMTVKGPGGNGAALVQDTASNSVNTQV 120
QY 141 GFGNNATRAQY 151
DB 121 GFGNNATAHQY 131

```

RESULT 3

```

; US-10-045-674A-594
; Sequence 594, Application US/10045674A
; GENERAL INFORMATION:
; APPLICANT: LADNER, ROBERT C.
; APPLICANT: COHEN, EDWARD H.
; APPLICANT: NASTRI, HORACIO G.
; APPLICANT: ROOKEY, KRISTIN L.
; APPLICANT: HOET, RENE
; TITLE OF INVENTION: NOVEL METHODS OF CONSTRUCTING LIBRARIES COMPRISING
; TITLE OF INVENTION: DISPLAYED AND/OR EXPRESSED MEMBERS OF A DIVERSE FAMILY
; TITLE OF INVENTION: OF PEPTIDES, POLYPEPTIDES OR PROTEINS AND THE NOVEL
; FILE REFERENCE: DVA/002 CIP2
; CURRENT APPLICATION NUMBER: US/10/045,674A
; PRIOR FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: 06/198,069
; PRIOR FILING DATE: 2000-04-17
; PRIOR APPLICATION NUMBER: 09/837,306
; PRIOR FILING DATE: 2001-04-17
; NUMBER OF SEQ ID NOS: 635
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 594

```

LENGTH: 434

```

; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: M13-III
; US-10-045-674A-594

```

Query Match

```

Best Local Similarity 10.6%; Score 82.5; DB 6; Length 434;
Best Local Similarity 24.1%; Pred. No. 3.2;
Matches 33; Conservative 19; Mismatches 48; Indels 37; Gaps 6;

```

```

QY 15 SGSLAGVPPWGGGNNHNGGSSGP-DYDQVLT---RVYTHMAHALQSDARKSETT 69
DB 259 SGGSSEGGSGGSGGSEGGSGGSDPYEKVANKAMTENADENALQSDA-KGKLD 317
QY 70 ITQSGYGNADVQAGADNSTIELTNGFRNNATIDQWNAKNSDITVGYGNNALVNOT 129
DB 318 SVATDYGAID-----GFIGDVS---GLANGATGDFAGSNSQMA--- 355
QY 130 ASDSSVWVRQYGRGNA 146
DB 356 -----QVGDGNS 363

```

RESULT 4

```

; US-10-045-674A-591
; Sequence 591, Application US/10045674A
; GENERAL INFORMATION:
; APPLICANT: LADNER, ROBERT C.
; APPLICANT: COHEN, EDWARD H.
; APPLICANT: NASTRI, HORACIO G.
; APPLICANT: ROOKEY, KRISTIN L.
; APPLICANT: HOET, RENE
; APPLICANT: HOOGENDOORN, HENDRICUS R. J. M.
; TITLE OF INVENTION: NOVEL METHODS OF CONSTRUCTING LIBRARIES COMPRISING
; TITLE OF INVENTION: DISPLAYED AND/OR EXPRESSED MEMBERS OF A DIVERSE FAMILY
; TITLE OF INVENTION: OF PEPTIDES, POLYPEPTIDES OR PROTEINS AND THE NOVEL
; FILE REFERENCE: DVA/002 CIP2
; CURRENT APPLICATION NUMBER: US/10/045,674A
; PRIOR FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: 06/198,069
; PRIOR FILING DATE: 2000-04-17
; PRIOR APPLICATION NUMBER: 09/837,306
; PRIOR FILING DATE: 2001-04-17
; NUMBER OF SEQ ID NOS: 635
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 591
; LENGTH: 424
; TYPE: PRT
; ORGANISM: Unknown Organism
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: M13 protein
; US-10-045-674A-591

```

Query Match

```

Best Local Similarity 10.4%; Score 81.5; DB 6; Length 424;
Best Local Similarity 25.2%; Pred. No. 3.8;
Matches 32; Conservative 19; Mismatches 37; Indels 39; Gaps 7;

```

```

QY 27 GGG--GNHNGGNSGP-DYDQVLT---RVYTHMAHALQSDARKSETTTTQSGYNGA 79
DB 259 GGSSEGGSGGSGGSGGSDPYEKVANKAMTENADENALQSDA-KGKLSVATDYGA 317
QY 80 DVGGADNSTIELTNGFRNNATIDQWNAKNSDITVQYGGNNALVNOTASDSSVWVRQ 139
DB 318 D-----GFIGDVS---GLANGATGDFAGSNSQMA-----Q 346
QY 140 VGFANNA 146
DB 347 VGDGNS 353

```

```

RESULT 5
US-10-045-674A-527
; Sequence 527, Application US/10045674A
; GENERAL INFORMATION:
; APPLICANT: LADNER, ROBERT C.
; APPLICANT: COHEN, EDWARD H.
; APPLICANT: NASTRI, HORACIO G.
; APPLICANT: ROOKEY, KRISTIN L.
; APPLICANT: HOET, RENE
; APPLICANT: HOOGENDOORN, HENDRICUS R. J. M.
; TITLE OF INVENTION: NOVEL METHODS OF CONSTRUCTING LIBRARIES COMPRISING
; TITLE OF INVENTION: DISPLAYED AND/OR EXPRESSED MEMBERS OF A DIVERSE FAMILY
; TITLE OF INVENTION: OF PEPTIDES, POLYPEPTIDES OR PROTEINS AND THE NOVEL
; FILE REFERENCE: DYAX/002 CIP2
; CURRENT APPLICATION NUMBER: US/10/045,674A
; PRIOR FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: 06/198,069
; PRIOR FILING DATE: 2000-04-17
; PRIOR APPLICATION NUMBER: 09/837,306
; PRIOR FILING DATE: 2001-04-17
; NUMBER OF SEQ ID NOS: 635
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 527
; LENGTH: 533
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Vector pCESS
US-10-045-674A-527

Query Match          10.4%; Score 81.5; DB 6; Length 533;
Best Local Similarity 25.2%; Pred. No. 5.1;
Matches 32; Conservative 19; Mismatches 37; Indels 39; Gaps 7;

QY 27 GGG--GNHGGGNSGSP--DYDQVLT-----RVYTHEMAHALQSDARKSETTITGSGYGNGA 79
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 368 GGGSEGGGSGGSGSDPDYEKMANANKAMTENADENALQSDA--KGKLDVSATDYGAAI 426
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 80 DVGQGANSTIELTONGFRNNATIDQMAKNSDITVGYGNNALVNOTASDSSVMWRO 139
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 427 D-----GFGDVS-----GLANGATGDPAGNSQMA-----Q 455
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 140 VGFGNNA 146
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 456 VGDGDS 462

RESULT 6
US-10-093-037A-63
; Sequence 63, Application US/10093037A
; GENERAL INFORMATION:
; APPLICANT: Jay M. Short
; APPLICANT: Bylina, Edward
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Mathur, Eric J.
; APPLICANT: Lam, David B.
; TITLE OF INVENTION: ENZYMES HAVING GLYCOSIDASE ACTIVITY AND METHODS OF USE THEREOF
; FILE REFERENCE: 56462001402
; CURRENT APPLICATION NUMBER: US/10/093,037A
; PRIOR FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: US 09/910,579
; PRIOR FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: US 09/134,078
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 08/949,026
; PRIOR FILING DATE: 1997-10-10
; PRIOR APPLICATION NUMBER: US 60/056,916
; PRIOR FILING DATE: 1996-12-06
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0

```

```

; SEQ ID NO 63
; LENGTH: 956
; TYPE: PRT
; ORGANISM: Bankia gouldi
US-10-093-037A-63

Query Match          10.2%; Score 79.5; DB 6; Length 956;
Best Local Similarity 22.1%; Pred. No. 16;
Matches 38; Conservative 18; Mismatches 67; Indels 49; Gaps 7;

QY 12 IVSSGALAGVPOWGGGNNHGGGNSGSDPDQVLT RVVTHEMAHALQSDARKSETTIT 71
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 318 LTAGSEIVNITIQNW-----DTEITGPK-----TTQCTIECTRAAETAQAGDEIIT 366
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 72 QSGYGAGDVGGADNSTIEL--TONGFRNNATI----- 103
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 367 APGNVNFQDKIGAFRRSYLYGSANNGSTNIIILGESATNPVPSGLDYNNGYLLSIE 426
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 104 -DQMAKNSDITVGYG-----GNNALVNOTASD---SSVMWROVGFGNNA 146
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 427 GDVWNIKDIFFYTGSGKIVLDNSNGSKLNLVVDHIGEEAHLRD--GSSNNS 477
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

RESULT 7
PCT-US04-05654-1498
; Sequence 1498, Application PC/TUS0405654
; GENERAL INFORMATION:
; APPLICANT: Sherman, Bradley K
; APPLICANT: Riechmann, Jose Luis
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Heard, Jacqueline E
; APPLICANT: Haake, Volker
; APPLICANT: Creelman, Robert A
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Adam, Luc J
; APPLICANT: Reuber, T. Lynne
; APPLICANT: Keddie, James
; APPLICANT: Dubell III, Arnold N
; APPLICANT: Pineda, Omaira
; APPLICANT: Repetti, Peter
; APPLICANT: Century, Karen
; APPLICANT: Guttersen, Neal
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Brown, Pierre E
; APPLICANT: Kumimoto, Roderick W
; APPLICANT: Pilgrim, Marsha L
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES IN PLANTS
; FILE REFERENCE: MBI-0047 PCT
; CURRENT APPLICATION NUMBER: PCT/US04/05654
; CURRENT FILING DATE: 2004-03-03
; PRIOR APPLICATION NUMBER: 10/374,780
; PRIOR FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: 10/675,852
; PRIOR FILING DATE: 2003-09-30
; NUMBER OF SEQ ID NOS: 2950
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1498
; LENGTH: 480
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Orthologous to G1198
PCT-US04-05654-1498

Query Match          10.1%; Score 78.5; DB 1; Length 480;
Best Local Similarity 26.8%; Pred. No. 8.3;
Matches 30; Conservative 19; Mismatches 54; Indels 9; Gaps 5;

QY 26 WGGGNNHNGGNSGSDPDQVLT RVVTHEMAHALQSDARKSETTITGSGYGAGVGGGA 85
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 45 WPDTSQGGGGGGAAYFGLEALVHVYA--TLRRRAQQTAT--TTSHGHTTPSTVA 102
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY 86 DNSTIELTONGFRNNATID--QMAKNSDITVGYGNNALVNOTASDSS 134
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

PRIOR FILING DATE: 1997-05-23

OTHER INFORMATION: clone ID: SORBI-28MAY03-C65560 1.pcp

APPLICANT: SHIMIZU, MASATOSHI

APPLICANT: SHIMIZU, MASATOSHI

APPLICANT: ITO, NOBUYA
 APPLICANT: WAKITA, RYUHEI
 TITLE OF INVENTION: PROCESS FOR PRODUCING OPTICALLY ACTIVE
 FILE REFERENCE: 4-HALO-3-HYDROXYBUTANOLATE
 CURRENT APPLICATION NUMBER: US/10/004,115B
 PRIOR FILING DATE: 2001-12-06
 PRIOR APPLICATION NUMBER: JP 2000-372704
 PRIOR FILING DATE: 2000-12-07
 PRIOR APPLICATION NUMBER: JP 2001-006144
 PRIOR FILING DATE: 2001-01-15
 PRIOR APPLICATION NUMBER: JP 2001-026594
 PRIOR FILING DATE: 2001-02-02
 PRIOR APPLICATION NUMBER: JP 2001-175175
 PRIOR FILING DATE: 2001-06-11
 NUMBER OF SEQ ID NOS: 37
 SOFTWARE: PatentIn Ver. 3.2
 SEQ ID NO 34
 LENGTH: 348
 TYPE: PRT
 ORGANISM: Corynebacterium sp.
 US-10-004-115B-34

Query Match 9.8%; Score 76.5; DB 6; Length 348;
 Best Local Similarity 24.1%; Pred. No. 8.4;
 Matches 33; Conservative 20; Mismatches 39; Indels 45; Gaps 6;

QY 45 QLVTRVVTHEMAHALQSDARKSETTITOS-----GY-----GNGADV-- 81
 DB 210 ELATKGAHEVLVS-DKDAENVRKTTGSGALVLPFGYQPTITMAVAGVSDVIT 268
 QY 82 ---GQADNSTIELTONGFRNNATTIDQWAKNS-----DITVGYGNNAA- 124
 DB 269 VGIQDGAHAKVGFQSPYEASVTVYWGARNELIELIDLAGIFDISVETSLDNGAE 328
 QY 125 ----LVNOTASDSSVAV 137
 DB 329 AYRRLAAGTISGRAVVV 345

RESULT 12
 US-10-451-467A-314
 Sequence 314, Application US/10451467A
 GENERAL INFORMATION:
 APPLICANT: CONTRERAS, ROLAND HENRI
 APPLICANT: EBERHARDT, INES
 APPLICANT: REEKANS, RIEKA JOSEPHINA
 APPLICANT: LUYTEN, WALTER HERMAN MARIA LOUIS
 TITLE OF INVENTION: BAX-RESPONSIVE GENES FOR DRUG TARGET IDENTIFICATION IN
 FILE REFERENCE: YEA-1667
 CURRENT APPLICATION NUMBER: US/10/451,467A
 PRIOR FILING DATE: 2003-06-19
 PRIOR APPLICATION NUMBER: EP 00870318.3
 PRIOR FILING DATE: 2000-12-22
 PRIOR APPLICATION NUMBER: EP 01870002.1
 PRIOR FILING DATE: 2001-01-04
 PRIOR APPLICATION NUMBER: EP 01870003.9
 PRIOR FILING DATE: 2001-01-09
 NUMBER OF SEQ ID NOS: 732
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO 314
 LENGTH: 430
 TYPE: PRT
 ORGANISM: Saccharomyces cerevisiae
 US-10-451-467A-314

Query Match 9.6%; Score 74.5; DB 6; Length 430;
 Best Local Similarity 21.1%; Pred. No. 17;
 Matches 27; Conservative 23; Mismatches 57; Indels 21; Gaps 5;
 QY 26 WGGGCHNGGNGSSGPDYDQVTRVVTHEMAHALQSDARKSETTITOSGNGADVGGGA 85

DB 138 YGSSNNNDYGSNNNDYS-----NNNDYGSNNNDYGSNNKKSYSGNNNDYS 191
 QY 86 DN-----STIELTONGFRNNATTIDQWAKNSDITVGYGNNAAVNOTASDS--SVAV 137
 DB 192 NNNDYGSNNKKSYSGSSNN--DSYGSNNND-----SYGSNNNDYGSNNNDYSNNK 244
 QY 138 ROYRGNN 145
 DB 245 KKSYSGSN 252

RESULT 13
 PCT-US04-05654-1406
 Sequence 1406, Application PC/TUS0405654
 GENERAL INFORMATION:
 APPLICANT: Sherman, Bradley K
 APPLICANT: Riechmann, Jose Luis
 APPLICANT: Jiang, Cai-Zhong
 APPLICANT: Heard, Jacqueline E
 APPLICANT: Haake, Volker
 APPLICANT: Creelman, Robert A
 APPLICANT: Ratcliffe, Oliver
 APPLICANT: Adam, Luc J
 APPLICANT: Reuber, T. Lynne
 APPLICANT: Reddie, James
 APPLICANT: Dubell III, Arnold N
 APPLICANT: Pineda, Omaira
 APPLICANT: Repetti, Peter
 APPLICANT: Century, Karen
 APPLICANT: Guterson, Neal
 APPLICANT: Yu, Guo-Liang
 APPLICANT: Brown, Pierre E
 APPLICANT: Kumimoto, Roderick W
 APPLICANT: Pilgrim, Marsha L
 TITLE OF INVENTION: POLYNICLOTIDES AND POLYPEPTIDES IN PLANTS
 FILE REFERENCE: MBI-0047 PCT
 CURRENT APPLICATION NUMBER: PCT/US04/05654
 PRIOR FILING DATE: 2004-03-03
 PRIOR APPLICATION NUMBER: 10/374,780
 PRIOR FILING DATE: 2003-02-25
 PRIOR APPLICATION NUMBER: 10/675,852
 PRIOR FILING DATE: 2003-09-30
 NUMBER OF SEQ ID NOS: 2950
 SOFTWARE: PatentIn version 3.2
 SEQ ID NO 1406
 LENGTH: 305
 TYPE: PRT
 ORGANISM: Oryza sativa
 FEATURE:
 OTHER INFORMATION: Orthologous to G1073
 PCT-US04-05654-1406

Query Match 9.5%; Score 74; DB 1; Length 305;
 Best Local Similarity 24.0%; Pred. No. 12;

Matches 29; Conservative 15; Mismatches 29; Indels 48; Gaps 5;
 QY 27 GGGN-----HNGGNSGPDYD-----QVTR-----VTHE 54
 DB 45 GGGNSNGPYPDHDGSSSPATDGAVGPGDVARRPRGRPSKPKKPPVITTE 104
 QY 55 MAHALQSDARKSETTITOSGNGADV-----GQADNSTIELTONGFRNNATTIDQW 107
 DB 105 SANTLRAHLE-----VSGCDVFECSVYARRRGVCSVLSGSGVTVTLRPS 155
 QY 108 A 108
 DB 156 A 156

RESULT 14
 PCT-US04-05654-1412
 Sequence 1412, Application PC/TUS0405654
 GENERAL INFORMATION:

```

: APPLICANT: Sherman, Bradley K
: APPLICANT: Riechmann, Jose Luis
: APPLICANT: Jiang, Cai-Zhong
: APPLICANT: Heard, Jacqueline E
: APPLICANT: Haake, Volker
: APPLICANT: Creelman, Robert A
: APPLICANT: Ratcliffe, Oliver
: APPLICANT: Adam, Luc J
: APPLICANT: Reuber, T. Lynne
: APPLICANT: Keddie, James
: APPLICANT: Dubell III, Arnold N
: APPLICANT: Pineda, Omaira
: APPLICANT: Repetti, Peter
: APPLICANT: Century, Karen
: APPLICANT: Guttersen, Neal
: APPLICANT: Yu, Guo-liang
: APPLICANT: Broun, Pierre E
: APPLICANT: Kumamoto, Roderick W
: APPLICANT: Pilgrim, Marsha L
: TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES IN PLANTS
: FILE REFERENCE: MBI-0047 PCT
: CURRENT APPLICATION NUMBER: PCT/US04/05654
: CURRENT FILING DATE: 2004-03-03
: PRIOR APPLICATION NUMBER: 10/374,780
: PRIOR FILING DATE: 2003-02-25
: PRIOR APPLICATION NUMBER: 10/675,852
: PRIOR FILING DATE: 2003-09-30
: NUMBER OF SEQ ID NOS: 2950
: SOFTWARE: PatentIn version 3.2
: SEQ ID NO 1412
: LENGTH: 310
: TYPE: PRT
: ORGANISM: Glycine max
: FEATURE:
: OTHER INFORMATION: Orthologous to G1073
PCT-US04-05654-1412

```

[illegible]

```
; TYPE: PRT
; ORGANISM: Sorghum bicolor
FEATURE:
; OTHER INFORMATION: clone ID: SORBI-28MAI03-C47268_1.pep
US-10-167-701-42417

Query Match          9.4%; Score 73; DB 6; Length 386;
Best Local Similarity 22.7%; Pred. No. 20;
Matches 27; Conservative 24; Mismatches 46; Indels 22; Gaps 5,

Oy      40  GPDYDLVTRVTYHEMAHALQSDARKSETTTTGGYGNGAGLVGQ----GANSTIELTON   95
        |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db       87  GTGSDQAVSR-----H-QPELKRITTTTTTGKHQTGLSGEELLIPDLSNVNISS   137
        |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Oy      96  GFRNNATTDWANNANSDITVGQYCGNNAALVNTPADSSVY-----VRÖVGFGNATA   148
        |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db     138  ---NSSPTPHNRKRVFDTLESSRGSQKVTVQETFSDDNKRPKPRPNRSQGEIGISHSAA  193
        |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
```

Search completed: March 11, 2004, 19:14:17
Job time : 6.3 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Comugen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:33:51 / Search time 171.3 Seconds
(without alignments)
860.386 Million cell updates/sec

Title: US-09-543-407-20

Perfect score: 774
Sequence: 1 MKLKVAAFAAIVSGSALA.....DSVWVROYFGNATANY 151

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 6019581 seqs, 976053577 residues

Total number of hits satisfying chosen parameters: 6019581

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents_AA_Main:*

1: /cgn2_6/ptodata/2/paa/US087_COMB.pep.*
2: /cgn2_6/ptodata/2/paa/US086_COMB.pep.*
3: /cgn2_6/ptodata/2/paa/US087_COMB.pep.*
4: /cgn2_6/ptodata/2/paa/US086_COMB.pep.*
5: /cgn2_6/ptodata/2/paa/US081_COMB.pep.*
6: /cgn2_6/ptodata/2/paa/US082_COMB.pep.*
7: /cgn2_6/ptodata/2/paa/US083_COMB.pep.*
8: /cgn2_6/ptodata/2/paa/US084_COMB.pep.*
9: /cgn2_6/ptodata/2/paa/US085_COMB.pep.*
10: /cgn2_6/ptodata/2/paa/US086_COMB.pep.*
11: /cgn2_6/ptodata/2/paa/US087_COMB.pep.*
12: /cgn2_6/ptodata/2/paa/US088_COMB.pep.*
13: /cgn2_6/ptodata/2/paa/US089_COMB.pep.*
14: /cgn2_6/ptodata/2/paa/US090_COMB.pep.*
15: /cgn2_6/ptodata/2/paa/US091_COMB.pep.*
16: /cgn2_6/ptodata/2/paa/US092_COMB.pep.*
17: /cgn2_6/ptodata/2/paa/US093_COMB.pep.*
18: /cgn2_6/ptodata/2/paa/US094_COMB.pep.*
19: /cgn2_6/ptodata/2/paa/US095_COMB.pep.*
20: /cgn2_6/ptodata/2/paa/US096_COMB.pep.*
21: /cgn2_6/ptodata/2/paa/US097A_COMB.pep.*
22: /cgn2_6/ptodata/2/paa/US097B_COMB.pep.*
23: /cgn2_6/ptodata/2/paa/US098_COMB.pep.*
24: /cgn2_6/ptodata/2/paa/US099A_COMB.pep.*
25: /cgn2_6/ptodata/2/paa/US099B_COMB.pep.*
26: /cgn2_6/ptodata/2/paa/US100_COMB.pep.*
27: /cgn2_6/ptodata/2/paa/US101_COMB.pep.*
28: /cgn2_6/ptodata/2/paa/US102_COMB.pep.*
29: /cgn2_6/ptodata/2/paa/US103_COMB.pep.*
30: /cgn2_6/ptodata/2/paa/US104_COMB.pep.*
31: /cgn2_6/ptodata/2/paa/US106_COMB.pep.*
32: /cgn2_6/ptodata/2/paa/US107_COMB.pep.*
33: /cgn2_6/ptodata/2/paa/US60_COMB.pep.*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	774	100.0	151 19	US-09-543-407-20
2	709	91.6	151 19	US-09-543-407-26
3	689	89.0	151 19	US-09-543-407-5
4	684	88.4	151 6	US-08-233-642A-57
5	662	85.5	151 19	US-09-543-407-28
6	622	80.4	151 19	US-09-543-407-18
7	619	80.0	151 19	US-09-543-407-22
8	611	78.9	151 19	US-09-543-407-12
9	609	78.7	151 19	US-09-543-407-14
10	606	78.3	151 19	US-09-543-407-24
11	602	77.8	131 19	US-09-543-407-11
12	597	77.1	151 19	US-09-543-407-30
13	574	74.2	151 19	US-09-543-407-16
14	528	68.2	151 19	US-09-543-407-7
15	525	67.8	151 13	US-08-978-878-4
16	525	67.8	151 21	US-09-741-873B-4
17	523	67.6	151 33	US-60-352-946-2
18	523	67.6	151 33	US-60-444-371-2
19	504	65.1	120 6	US-08-233-642A-55
20	470	60.7	109 19	US-09-543-407-34
21	466	60.2	158 16	US-09-252-691C-5834
22	466	60.2	158 16	US-09-252-691C-5834
23	466	60.2	158 30	US-10-417-886-5834
24	447	57.8	131 13	US-08-978-878-2
25	447	57.8	131 21	US-09-741-873B-2
26	343	44.3	109 19	US-09-543-407-35
27	263	34.0	68 19	US-09-543-407-37
28	166.5	21.5	70 19	US-09-543-407-32
29	153	19.8	48 19	US-09-543-407-19
30	105.5	13.6	145 21	US-09-733-449-8854
31	105.5	13.6	145 23	US-09-803-110-8854
32	101.5	11.1	445 29	US-10-369-493-20638
33	101.5	11.1	445 33	US-60-369-039-20638
34	99	12.8	492 21	US-09-708-427-7657
35	99	12.8	544 21	US-09-708-427-7656
36	99	12.8	573 21	US-09-708-427-7655
37	99	12.8	955 24	US-09-935-625-7946
38	99	12.8	955 24	US-09-935-625-7945
39	99	12.8	1036 24	US-09-935-625-7945
40	99	12.8	1036 24	US-09-935-625-7945
41	99	12.8	1055 19	US-09-570-581A-1851
42	99	12.8	1055 19	US-09-573-655A-349
43	99	12.8	1055 19	US-09-573-655A-667
44	99	12.8	1055 19	US-09-573-655B-349
45	99	12.8	1055 19	US-09-573-655B-667

ALIGNMENTS

RESULT 1
US-09-543-407-20
Sequence 20, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIBRILLAR SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543,407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 20
LENGTH: 151
TYPE: PRT
ORGANISM: Artificial Sequence
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.

US-09-543-407-20

Query Match 100.0%; Score 774; DB 19; Length 151;
 Best Local Similarity 100.0%; Pred. No. 3.3e-75;
 Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLKVAFAFAIIVSGSALAGVVPQWGGGNNHGGSSGPDSTLSIYOGSANAALALQ 60
 DB 1 MKLKVAFAFAIIVSGSALAGVVPQWGGGNNHGGSSGPDSTLSIYOGSANAALALQ 60
 QY 61 SDARKYDQVTRVYTHMAHAGGADNSTIELTONGFRNNATIDQWNAKNSDITVGOYGG 120
 DB 61 SDARKYDQVTRVYTHMAHAGGADNSTIELTONGFRNNATIDQWNAKNSDITVGOYGG 120
 QY 121 NNAALVNOTASDSSVWVROVGFNNATANQY 151
 DB 121 NNAALVNOTASDSSVWVROVGFNNATANQY 151

RESULT 2

US-09-543-407-26

Sequence 26, Application US/09543407

GENERAL INFORMATION:
 APPLICANT: White, Aaron P.
 APPLICANT: Doran, James L.
 APPLICANT: Collinson, S. Karen
 APPLICANT: Kay, William W.
 TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
 FILE REFERENCE: 920043.406
 CURRENT APPLICATION NUMBER: US/09/543.407
 CURRENT FILING DATE: 2000-04-05
 NUMBER OF SEQ ID NOS: 59
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 26
 LENGTH: 151
 TYPE: PRT
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afga
 OTHER INFORMATION: sequence containing the replacement fragment
 OTHER INFORMATION: encoding P73 from GPe3 of *Leishmania major*.
 US-09-543-407-26

Query Match 91.6%; Score 709; DB 19; Length 151;
 Best Local Similarity 89.9%; Pred. No. 3.8e-68;
 Matches 143; Conservative 0; Mismatches 0; Indels 16; Gaps 2;

QY 1 MKLKVAFAFAIIVSGSALAGVVPQWGGGNNHGGSSGPDSTLSIYOGSANAALALQ 60
 DB 1 MKLKVAFAFAIIVSGSALAGVVPQWGGGNNHGGSSGPDSTLSIYOGSANAALALQ 57
 QY 61 SDARKYDQVTRVYTHMAHAGGADNSTIELTONGFRNNATIDQWNAKNSD 112
 DB 58 -----YDQVTRVYTHMAHAGGADNSTIELTONGFRNNATIDQWNAKNSD 112
 QY 113 ITVGOYGNNAAALVNOTASDSSVWVROVGFNNATANQY 151
 DB 113 ITVGOYGNNAAALVNOTASDSSVWVROVGFNNATANQY 151

RESULT 3

US-09-543-407-5

Sequence 5, Application US/09543407

GENERAL INFORMATION:
 APPLICANT: White, Aaron P.
 APPLICANT: Doran, James L.
 APPLICANT: Collinson, S. Karen
 APPLICANT: Kay, William W.
 TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
 FILE REFERENCE: 920043.406
 CURRENT APPLICATION NUMBER: US/09/543.407

CURRENT FILING DATE: 2000-04-05
 NUMBER OF SEQ ID NOS: 59
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 5
 LENGTH: 151
 TYPE: PRT
 ORGANISM: *Salmonella enteritidis*
 US-09-543-407-5

Query Match 89.0%; Score 689; DB 19; Length 151;
 Best Local Similarity 90.7%; Pred. No. 5.6e-66;
 Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 1 MKLKVAFAFAIIVSGSALAGVVPQWGGGNNHGGSSGPDSTLSIYOGSANAALALQ 60
 DB 1 MKLKVAFAFAIIVSGSALAGVVPQWGGGNNHGGSSGPDSTLSIYOGSANAALALQ 60
 QY 61 SDARKYDQVTRVYTHMAHAGGADNSTIELTONGFRNNATIDQWNAKNSDITVGOYGG 120
 DB 61 SDARKSETTITGCGYNGADVGGADNSTIELTONGFRNNATIDQWNAKNSDITVGOYGG 120
 QY 121 NNAALVNOTASDSSVWVROVGFNNATANQY 151
 DB 121 NNAALVNOTASDSSVWVROVGFNNATANQY 151

RESULT 4

US-08-233-642A-57

Sequence 57, Application US/08233642A

GENERAL INFORMATION:
 APPLICANT: Kay, William W.
 APPLICANT: Collinson, S. Karen
 APPLICANT: Clouthier, Sharon C.
 APPLICANT: Doran, James L.
 TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR *SALMONELLA*-
 NUMBER OF SEQUENCES: 58
 NUMBER OF SEQUENCES: -
 CORRESPONDENCE ADDRESS:
 ADDRESSER: Seed and Berry
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington
 COUNTRY: U.S.A.
 ZIP: 98104-7092
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/233,642A
 FILING DATE: 26-APR-1994
 CLASSIFICATION: 424
 ATTORNEY/AGENT INFORMATION:
 NAME: King, Joshua
 REGISTRATION NUMBER: 35,570
 REFERENCE/DOCKET NUMBER: 920043.403C3
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 622-4900
 TELEFAX: (206) 682-6031
 TELEX: 3723836 SEDANBERRY
 INFORMATION FOR SEQ ID NO: 57:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 151 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-233-642A-57

Query Match 88.4%; Score 684; DB 6; Length 151;
 Best Local Similarity 90.1%; Pred. No. 1.9e-65;
 Matches 136; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

```
QY      1 MLLKVAAPFAIYVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYOGSANAALALQ 60
      1 MLLKVAAPFAIYVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYOGSANAALALQ 60
Db      61 SPARKDQDLVTRVVTHEMAHAGGADNSTIELTQNGFRNNATIDQWNAKSDITVGOYGG 120
      61 SPARKSETTITGSGYNGADVDGQGANSTIELTQNGFRNNATIDQWNAKSDITVGOYGG 120
Db      121 NNAALVNOTASDSSVMVROVFGNNATANOY 151
      121 NNPALVNOTASDSSVMVROVFGNNATANOY 151
Db

RESULT 5
US-09-543-407-28
; Sequence 28, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-28

Query Match      85.5%; Score 662; DB 19; Length 151;
Best Local Similarity 81.9%; Pred. No. 4.7e-63;
Matches 136; Conservative 0; Mismatches 0; Indels 30; Gaps 2;

QY      1 MLLKVAAPFAIYVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYOGSANAALALQ 60
      1 MLLKVAAPFAIYVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYOGSANAALALQ 60
Db      61 SPARKDQDLVTRVVTHEMAHAGGADNSTIELTQNGFRNNATIDQ 105
      61 SPARKSETTITGSGYNGADVDQDLVTRVVTHEMAHA-----FRNNATIDQ 105
Db      106 WNAKNSDITVGOYGNNALVNOTASDSSVMVROVFGNNATANOY 151
      106 WNAKNSDITVGOYGNNALVNOTASDSSVMVROVFGNNATANOY 151
Db

RESULT 6
US-09-543-407-18
; Sequence 18, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 151
```

```
      ; TYPE: PRT
      ; ORGANISM: Artificial Sequence
      ; FEATURE:
      ; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
      ; OTHER INFORMATION: sequence containing the replacement fragment
      ; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-18

Query Match      80.4%; Score 622; DB 19; Length 151;
Best Local Similarity 73.6%; Pred. No. 1e-58;
Matches 128; Conservative 0; Mismatches 0; Indels 46; Gaps 2;

QY      1 MLLKVAAPFAIYVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYOGSANAALALQ 60
      1 MLLKVAAPFAIYVSGSALAGVVPQWGGGNNHGGNSGPD-----
Db      61 SPARKDQDLVTRVVTHEMAHAGGADNSTIELTQNGF 97
      61 SPARKDQDLVTRVVTHEMAHAGGADNSTIELTQNGF 97
Db      43 -----YDQLVTRVVTHEMAHAGGADNSTIELTQNGF 97
      43 -----YDQLVTRVVTHEMAHAGGADNSTIELTQNGF 97
QY      98 RNNATIDQWNAKNSDITVGOYGNNALVNOTASDSSVMVROVFGNNATANOY 151
      98 RNNATIDQWNAKNSDITVGOYGNNALVNOTASDSSVMVROVFGNNATANOY 151
Db      98 RNNATIDQWNAKNSDITVGOYGNNALVNOTASDSSVMVROVFGNNATANOY 151

RESULT 7
US-09-543-407-22
; Sequence 22, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 22
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-22

Query Match      80.0%; Score 619; DB 19; Length 151;
Best Local Similarity 74.6%; Pred. No. 2.2e-58;
Matches 129; Conservative 0; Mismatches 0; Indels 44; Gaps 2;

QY      1 MLLKVAAPFAIYVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYOGSANAALALQ 60
      1 MLLKVAAPFAIYVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYOGSANAALALQ 60
Db      61 SPARKDQDLVTRVVTHEMAHAGGADNSTIELTQNGR 98
      61 SPARKSETTITGSGYNGADVDQGANVDQDLVTRVVTHEMAHA----- 103
Db      99 NNAATIDQWNAKNSDITVGOYGNNALVNOTASDSSVMVROVFGNNATANOY 151
      104 -----DQWNAKNSDITVGOYGNNALVNOTASDSSVMVROVFGNNATANOY 151
Db

RESULT 8
US-09-543-407-12
; Sequence 12, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
```

APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIBRILLAR SYSTEM FOR
TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543,407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12
LENGTH: 151
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-12

Query Match 78.9%; Score 611; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 1.6e-57;
Matches 122; Conservative 7; Mismatches 22; Indels 0; Gaps 0;

QY 1 MLLKVAAPFAIVSGSALAGVVPWGGGNNHGGNSGPDSTLSIYGSANALALQ 60
DB 1 MLLKVAAPFAIVSGSALAGVVPWGGGNNHGGNSGPDSTLSIYGSANALALQ 60
QY 61 SDARKYDQLVTRVVTHEMAHAGGADNSTIELTONGFNNATTDQWAKNSDITVGYGG 120
DB 61 SDARKSETTITGSGYGADVDGQADNSTIELTONGFNNATTDQWAKNSDITVGYGG 120
QY 121 NNAALVNOTASDSSVMRVQVFGNNATANY 151
DB 121 NNAALVNYDQLVTRVVTHEMAHANNATANY 151

RESULT 9
US-09-543-407-14
Sequence 14, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIBRILLAR SYSTEM FOR
TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543,407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 14
LENGTH: 151
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-14

Query Match 78.7%; Score 609; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 2.7e-57;
Matches 123; Conservative 6; Mismatches 22; Indels 0; Gaps 0;

QY 1 MLLKVAAPFAIVSGSALAGVVPWGGGNNHGGNSGPDSTLSIYGSANALALQ 60
DB 1 MLLKVAAPFAIVSGSALAGVVPWGGGNNHGGNSGPDSTLSIYGSANALALQ 60
QY 61 SDARKYDQLVTRVVTHEMAHAGGADNSTIELTONGFNNATTDQWAKNSDITVGYGG 120
DB 61 SDARKSETTITGSGYGADVDGQADNSTIELTONGFNNATTDQWAKNSDITVGYGG 120

QY 121 NNAALVNOTASDSSVMRVQVFGNNATANY 151
DB 121 LVTRVVTHEMAHASVMRVQVFGNNATANY 151

RESULT 10
US-09-543-407-24
Sequence 24, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIBRILLAR SYSTEM FOR
TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543,407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 24
LENGTH: 151
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-24

Query Match 78.3%; Score 606; DB 19; Length 151;
Best Local Similarity 82.1%; Pred. No. 5.7e-57;
Matches 124; Conservative 5; Mismatches 22; Indels 0; Gaps 0;

QY 1 MLLKVAAPFAIVSGSALAGVVPWGGGNNHGGNSGPDSTLSIYGSANALALQ 60
DB 1 MLLKVAAPFAIVSGSALAGVVPWGGGNNHGGNSGPDSTLSIYGSANALALQ 60
QY 61 SDARKYDQLVTRVVTHEMAHAGGADNSTIELTONGFNNATTDQWAKNSDITVGYGG 120
DB 61 SDARKSETTITGSGYGADVDGQADNSTIELTONGFNNATTDQWAKNSDITVGYGG 120
QY 121 NNAALVNOTASDSSVMRVQVFGNNATANY 151
DB 121 HEMAANOTASDSSVMRVQVFGNNATANY 151

RESULT 11
US-09-543-407-31
Sequence 31, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIBRILLAR SYSTEM FOR
TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543,407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 31
LENGTH: 131
TYPE: PRT
ORGANISM: Salmonella enteritidis
US-09-543-407-31

Query Match 77.8%; Score 602; DB 19; Length 131;
Best Local Similarity 89.3%; Pred. No. 1.3e-56;
Matches 117; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 21 GVPWGGGNNHGGNSGPDSTLSIYGSANALALQSDARKYDQLVTRVVTHEMAH 80

```
Db 1 GVPWPQGGGNNHGGSSGPDSTLSIYOGSANAALALQSDARKSETTITGSGYNGAD 60
QY 81 AAGGADNSTIELTONGFRNNATIDQNNAKNSDITVGOYGGNNAALVNOTASDSVWVROV 140
Db 61 VGGADNSTIELTONGFRNNATIDQNNAKNSDITVGOYGGNNAALVNOTASDSVWVROV 120
QY 141 GFGNNATANQY 151
Db 121 GFGNNATANQY 131

RESULT 12
US-09-543-407-30
; Sequence 30, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GPe3 of Leishmania major.
US-09-543-407-30

Query Match 77.1%; Score 597; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 5.4e-56;
Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;

QY 1 MLKLKVAAPAAIYVSGSALAGVVPQWGGGNNHGGSSGPDSTLSIYOGSANAALALQ 60
Db 1 MLKLKVAAPAAIYVSGSALAGVVPQWGGGNNHGGSSGPDSTLSIYOGSANAALALQ 60
QY 61 SPARKYDQLVTRVVTHEMAAGGADNSTIELTONGFRNNATIDQNNAKNSDITVGOYGG 120
Db 61 SPARKSETTITGSGYNGADVGQGDNSTIELTONGFRNNATIDQLVTRVVTHEMAAGG 120
QY 121 NNAALVNOTASDSVWVROVGFNNATANQY 151
Db 121 NNAALVNOTASDSVWVROVGFNNATANQY 151

RESULT 13
US-09-543-407-16
; Sequence 16, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
```

```
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GPe3 of Leishmania major.
US-09-543-407-16

Query Match 74.2%; Score 574; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 1.7e-53;
Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;

QY 1 MLKLKVAAPAAIYVSGSALAGVVPQWGGGNNHGGSSGPDSTLSIYOGSANAALALQ 60
Db 1 MLKLKVAAPAAIYVSGSALAGVDPQVTRVVTHEMAHAGSPDSTLSIYOGSANAALALQ 60
QY 61 SPARKYDQLVTRVVTHEMAAGGADNSTIELTONGFRNNATIDQNNAKNSDITVGOYGG 120
Db 61 SPARKSETTITGSGYNGADVGQGDNSTIELTONGFRNNATIDQNNAKNSDITVGOYGG 120
QY 121 NNAALVNOTASDSVWVROVGFNNATANQY 151
Db 121 NNAALVNOTASDSVWVROVGFNNATANQY 151

RESULT 14
US-09-543-407-7
; Sequence 7, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-543-407-7

Query Match 68.2%; Score 528; DB 19; Length 151;
Best Local Similarity 68.9%; Pred. No. 1.6e-48;
Matches 104; Conservative 20; Mismatches 27; Indels 0; Gaps 0;

QY 1 MLKLKVAAPAAIYVSGSALAGVVPQWGGGNNHGGSSGPDSTLSIYOGSANAALALQ 60
Db 1 MLKLKVAAPAAIYVSGSALAGVVPQWGGGNNHGGSSGPNSEINITYOGGNSALALQ 60
QY 61 SPARKYDQLVTRVVTHEMAAGGADNSTIELTONGFRNNATIDQNNAKNSDITVGOYGG 120
Db 61 TDARNSDLTITGHHGGNGADVGQSDSDSIDLTGFGNSATLIDQNGHNSDITVQAFGG 120
QY 121 NNAALVNOTASDSVWVROVGFNNATANQY 151
Db 121 GNGAADVDTASNSVWVROVGFNNATANQY 151

RESULT 15
US-08-978-878-4
; Sequence 4, Application US/08978878
; GENERAL INFORMATION:
; APPLICANT: NORMARK, Steffan
; APPLICANT: OLSEN, Arne
; TITLE OF INVENTION: FIBRONECTIN BINDING PROTEIN AS WELL AS ITS PREPARATION
; FILE REFERENCE: 012889-081
; CURRENT APPLICATION NUMBER: US/08/978,878
; CURRENT FILING DATE: 1997-11-26
; EARLIER APPLICATION NUMBER: SE 8801723-1
; EARLIER FILING DATE: 1988-05-06
```

EARLIER APPLICATION NUMBER: US 07/347,189
EARLIER FILING DATE: 1989-05-04
EARLIER APPLICATION NUMBER: US 07/789,437
EARLIER FILING DATE: 1991-11-06
EARLIER APPLICATION NUMBER: US 07/970,846
EARLIER FILING DATE: 1992-11-03
EARLIER APPLICATION NUMBER: US 08/187,865
EARLIER FILING DATE: 1994-01-28
EARLIER APPLICATION NUMBER: US 08/318,519
EARLIER FILING DATE: 1994-10-05
EARLIER APPLICATION NUMBER: US 08/495,959
EARLIER FILING DATE: 1995-06-28
NUMBER OF SEQ ID NOS: 10
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO: 4
LENGTH: 151
TYPE: PRT
ORGANISM: Escherichia coli
US-08-978-878-4

Query Match 67.8%; Score 525; DB 13; Length 151;
Best Local Similarity 68.2%; Pred. No. 3.5e-48;
Matches 103; Conservative 21; Mismatches 27; Indels 0; Gaps 0;

QY 1 MKLKVAFPAIVSGSALAGVFPQWGGGNHNGGSSGPDSTLSTIYQGSANALALQ 60
DB 1 MKLKVAFPAIVSGSALAGVFPQWGGGNHNGGSSGPDSTLSTIYQGSANALALQ 60
QY 61 SDARKYDQLVTRVTHMAHAGGAGDNSTIELTONGFPNNATIDOWNAXNSDITVGOYG 120
DB 61 TDARNSDLITTOHGGGNCADYGGSSDSIDLITQRFNGSKATLDQWNGKNSMTVKQFGG 120
QY 121 NNALVNOTASDSSVMVRQVGFNNATANQY 151
DB 121 GNGAAVDQTAASNSVNTQVGFNNATAHQY 151

Search completed: March 11, 2004, 19:13:12
Job time : 172.3 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:35:05 ; Search time 5.3 Seconds
(without alignments)
376.014 Million cell updates/sec

Title: US-09-543-407-20

Sequence: 1 MLLKVAAPFAIIVSGSALA.....DSSVMROYFGNNTANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 125546 seqs, 13197846 residues

Total number of hits satisfying chosen parameters: 125546

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Pending Patents_AA_New.*

1: /cgn2_6/prodata/1/paa/PCT_NEW_COMB.pep.*
2: /cgn2_6/prodata/1/paa/US06_NEW_COMB.pep.*
3: /cgn2_6/prodata/1/paa/US07_NEW_COMB.pep.*
4: /cgn2_6/prodata/1/paa/US08_NEW_COMB.pep.*
5: /cgn2_6/prodata/1/paa/US09_NEW_COMB.pep.*
6: /cgn2_6/prodata/1/paa/US10_NEW_COMB.pep.*
7: /cgn2_6/prodata/1/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	525	67.8	151	US-09-741-873C-4	Sequence 4, Appl1
2	447	57.8	131	US-09-741-873C-2	Sequence 2, Appl1
3	76.5	9.9	179	US-10-767-701-35342	Sequence 35342, A
4	73.5	9.5	250	US-10-779-461-40	Sequence 40, Appl1
5	73.5	9.5	424	US-10-045-674A-591	Sequence 591, App
6	73.5	9.5	433	PCT-US04-05654-2586	Sequence 527, App
7	73.5	9.5	533	US-10-045-674A-527	Sequence 590, App
8	73	9.4	321	PCT-US04-05654-590	Sequence 594, App
9	72.5	9.4	434	US-10-045-674A-594	Sequence 594, App
10	72	9.3	708	US-10-417-884A-6047	Sequence 6047, App
11	72	9.3	816	PCT-US04-03808-14	Sequence 14, Appl1
12	71.5	9.2	956	US-10-093-037A-63	Sequence 63, Appl1
13	71.5	9.2	2364	US-09-126-816C-6	Sequence 6, Appl1
14	71	9.2	363	US-10-767-701-37968	Sequence 37968, A
15	71	9.2	399	PCT-US04-05654-2748	Sequence 2748, App
16	70.5	9.1	471	US-10-746-795A-22	Sequence 22, Appl1
17	69.5	9.0	1160	US-10-603-150-2	Sequence 2, Appl1
18	69	8.9	1327	PCT-US04-02338-49	Sequence 49, Appl1
19	68.5	8.9	574	PCT-US04-05654-334	Sequence 334, App
20	68.5	8.9	878	US-10-771-241-364	Sequence 364, App
21	68.5	8.9	1531	PCT-US03-32645-44	Sequence 44, Appl1
22	68	8.8	139	US-10-767-701-37368	Sequence 37368, A
23	68	8.8	351	PCT-US04-05654-1755	Sequence 1755, App
24	68	8.8	1172	US-10-451-467A-574	Sequence 574, App
25	67.5	8.7	359	US-10-767-701-40820	Sequence 40820, A
26	67	8.7	447	US-10-767-701-45219	Sequence 45219, A

27	66.5	8.6	348	US-10-004-115B-34	Sequence 34, Appl1
28	66.5	8.6	362	US-10-767-701-37331	Sequence 37331, A
29	66.5	8.6	624	PCT-US04-05654-1466	Sequence 1466, App
30	66.5	8.6	1290	US-10-786-892-138	Sequence 138, App
31	66	8.5	164	US-10-767-701-42650	Sequence 42650, A
32	66	8.5	411	US-10-767-701-43919	Sequence 43919, A
33	66	8.5	420	US-10-786-850-2	Sequence 2, Appl1
34	66	8.5	609	US-10-151-553-2	Sequence 2, Appl1
35	65.5	8.5	138	US-10-779-461-43	Sequence 48913, A
36	65.5	8.5	257	US-10-779-461-43	Sequence 43, Appl1
37	65.5	8.5	318	PCT-US04-05654-120	Sequence 5818, App
38	65.5	8.5	373	US-10-600-070-119	Sequence 119, App
39	65.5	8.5	566	US-10-695-499-70	Sequence 70, Appl1
40	65.5	8.5	619	US-10-417-884A-5134	Sequence 5134, App
41	65.5	8.5	673	US-10-600-070-169	Sequence 169, App
42	65.5	8.5	714	US-10-600-070-1170	Sequence 170, App
43	65.5	8.5	943	PCT-US04-05654-1838	Sequence 1838, App
44	65.5	8.5	150	US-10-767-701-59520	Sequence 59520, A
45	65	8.4	6	US-10-767-701-59520	Sequence 59520, A

ALIGNMENTS

RESULT 1
US-09-741-873C-4
; Sequence 4, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Steffen
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873C-4

Query Match 67.8%; Score 525; DB 5; Length 151;
Best Local Similarity 68.2%; Pred. No. 2.4e-41;
Matches 103; Conservative 21; Mismatches 27; Indels 0; Gaps 0;

QY	1	MLLKVAAPFAIIVSGSALAGVPPWGGGNGGNSGSPSTSIYIGSANALAIQ	60
DB	1	MLLKVAAPFAIIVFSGSAVAGVPPYGGGNGGNGGNSGSPSEINITYGGNSALAIQ	60
QY	61	SNARKYDOLVTVVTHMAHAGGAGDNSTIEHTONGFRNATIDQWNAKSDITVGGYGG	120
DB	61	TDAKSDLTITTHGGGNGAGADVQGSDDSSITDTQGFSGSATLDQWNGKNSERTVQFGG	120
QY	121	NNAALVNGTASDSVVRQVFGNNTANQY	151
DB	121	GNGAADVQTAASNSVVRQVFGNNTANQY	151

RESULT 2
US-09-741-873C-2
; Sequence 2, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibrinectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741, 873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978, 878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347, 189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789, 437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970, 846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187, 865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318, 519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873C-2

Query Match 57.8%; Score 447; DB 5; Length 131;
Best Local Similarity 64.9%; Pred. No. 2,7e-34;
Matches 85; Conservative 20; Mismatches 26; Indels 0; Gaps 0;

QY 21 GVVQWGGGNNHNGGNSGSPDSTLSTIYQGSANALALQSDARKYDQVTRVVTHEMAH 80
DB 1 GVVQYGGGNNHNGGNSGSPDSTLSTIYQGSANALALQSDARKYDQVTRVVTHEMAH 60
QY 81 AGGADNSTIELTONGFRNATIDOWNAKNSDIVQYGGNNALVNOTASDSSVMYROY 140
DB 61 VGGSDDBSIDLTRGFRNSATLIDQNGKNSSEMTVKQFGGNGAADVDTASSSVNTQY 120
QY 141 GFGNNATANOY 151
DB 121 GFGNNATANOY 131

RESULT 3
US-10-767-701-35342
; Sequence 35342, Application US/10767701
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767, 701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 35342
; LENGTH: 179
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-C65560_1.pcp
US-10-767-701-35342

Query Match 9.9%; Score 76.5; DB 6; Length 179;
Best Local Similarity 26.0%; Pred. No. 3.1;
Matches 33; Conservative 13; Mismatches 46; Indels 35; Gaps 6;

QY 27 GGGNNHNGGNSGSPDSTLSTIYQGSANALALQSDARKYDQVTRVVTHEMAHAGQAD 86
DB 4 GGRNRYSGGTSS-----YYEG-----RRRRRSYQSRVSNDFV-DAASTAVGAD 46
QY 87 NSTIELTONGFRNATIDOWNAKNSDIVQYGGNNALVNOTA-----SSSSVMYQ 139
DB 47 DNT-----SSSSGDDGWSKSS--GGYGWDEILLDNNTAAAOEAKRSSNSPSAGN 95
QY 140 VGFNNNA 146
DB 96 NGYGTSS 102

RESULT 4
US-10-779-461-40
; Sequence 40, Application US/10779461
; GENERAL INFORMATION:
; APPLICANT: Morton, Philip A
; TITLE OF INVENTION: ANTIBODIES TO C-MET FOR THE TREATMENT OF CANCERS
; FILE REFERENCE: 00980/1
; CURRENT APPLICATION NUMBER: US/10/779, 461
; CURRENT FILING DATE: 2004-02-13
; PRIOR APPLICATION NUMBER: 60/447, 073
; PRIOR FILING DATE: 2003-02-13
; NUMBER OF SEQ ID NOS: 161
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 40
; LENGTH: 250
; TYPE: PRT
; ORGANISM: artificial
; FEATURE:
; OTHER INFORMATION: phage display generated human antibody
US-10-779-461-40

Query Match 9.5%; Score 73.5; DB 6; Length 250;
Best Local Similarity 20.5%; Pred. No. 9;
Matches 32; Conservative 28; Mismatches 67; Indels 29; Gaps 5;

QY 9 FAATVSSGALAGVYPONG-----GGNHNHNGGNSGSPDSTLSTIYQGSAN----- 54
DB 95 YCAISPLRGLTADVDVWQGTLYVSSGGGSGGGSQASALTQPAASASGSPQ 154
QY 55 -----AALALQSDARKYDQV-----VTRVVTHEMAHAGQADNSTIELTONGFRNNA 101
DB 155 SITISCTGSSDIGRIDVSWTQRPKAPKLMYDIVINRPSGV-SSRFSKSGSNTASL 213
QY 102 TIDOWNAKN-SDITVQYGGNNALVNOTASDSSVM 136
DB 214 TISGLQAEDEADYCSYAGSTTLVFGTGTGLTVL 249

RESULT 5
US-10-045-674A-591
; Sequence 591, Application US/10045674A
; GENERAL INFORMATION:
; APPLICANT: LADNER, ROBERT C.
; APPLICANT: COHEN, EDWARD H.
; APPLICANT: NASTRI, HORACIO G.
; APPLICANT: ROONEY, KRISTIN L.
; APPLICANT: HOEFT, RENE
; TITLE OF INVENTION: NOVEL METHODS OF CONSTRUCTING LIBRARIES COMPRISING
; TITLE OF INVENTION: DISPLAYED AND/OR EXPRESSED MEMBERS OF A DIVERSE FAMILY
; TITLE OF INVENTION: OF PEPTIDES, POLYPEPTIDES OR PROTEINS AND THE NOVEL
; FILE REFERENCE: DIXA/002 C1P2
; CURRENT APPLICATION NUMBER: US/10/045, 674A
; CURRENT FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: 06/198, 069
; PRIOR FILING DATE: 2000-04-17
; PRIOR APPLICATION NUMBER: 09/837, 306
; PRIOR FILING DATE: 2001-04-17

NUMBER OF SEQ ID NOS: 635
 SOFTWARE: Patentin Ver. 2.1
 SEQ ID NO 591
 LENGTH: 424
 TYPE: PRT
 ORGANISM: Unknown Organism
 FEATURE:
 OTHER INFORMATION: Description of Unknown Organism: M13 protein
 US-10-045-674A-591

Query Match 9.5%; Score 73.5; DB 6; Length 424;
 Best Local Similarity 25.7%; Pred. No. 18;
 Matches 37; Conservative 15; Mismatches 47; Indels 45; Gaps 8;

QY 15 SCSALAGVVPWQCG--GGNHNCGNSGPDSTLSIYQ-YGSANAAL-----ALQSDA 63
 DB 243 SGGSGEGGSGEGGSGEGGSGEGGSGGSDPDEKMANNKAMTENADENALQSDA 302
 QY 64 R-KYDQLVTVRYVTHEMAHAGGADNSTIELTONGFPNNATTIDOMAKNSDITVGYQGN 122
 DB 303 KKKLDSVAT-----DYGAID-----GFIGDVS-----GLANGGATGDPAGSN 341
 QY 123 AALVNGTASDSSVMVROYFGNNA 146
 DB 342 SQMA-----QVGDGDS 353

RESULT 6
 PCT-US04-05654-2586
 Sequence 2586, Application PC/TUS0405654

GENERAL INFORMATION:
 APPLICANT: Sherman, Bradley K
 APPLICANT: Riechmann, Jose Luis
 APPLICANT: Jiang, Cai-Zhong
 APPLICANT: Haake, Jacqueline E
 APPLICANT: Haake, Volker
 APPLICANT: Creelman, Robert A
 APPLICANT: Ratcliffe, Oliver
 APPLICANT: Adam, Luc J
 APPLICANT: Reuber, T. Lynne
 APPLICANT: Keddie, James
 APPLICANT: Dubell III, Arnold N
 APPLICANT: Pineda, Omaira
 APPLICANT: Repetti, Peter
 APPLICANT: Century, Karen
 APPLICANT: Guterson, Neal
 APPLICANT: Yu, Guo-Liang
 APPLICANT: Broun, Pierre E
 APPLICANT: Kumimoto, Roderick W
 APPLICANT: Pilgrim, Marsha L
 TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES IN PLANTS
 FILE REFERENCE: MBI-0047 PCT
 CURRENT APPLICATION NUMBER: PCT/US04/05654
 CURRENT FILING DATE: 2004-03-03
 PRIOR APPLICATION NUMBER: 10/374,780
 PRIOR FILING DATE: 2003-02-25
 PRIOR APPLICATION NUMBER: 10/675,852
 PRIOR FILING DATE: 2003-09-30
 NUMBER OF SEQ ID NOS: 2950
 SOFTWARE: Patentin version 3.2
 SEQ ID NO 2586
 LENGTH: 433
 TYPE: PRT
 ORGANISM: Arabidopsis thaliana
 FEATURE:
 OTHER INFORMATION: G1337
 PCT-US04-05654-2586

Query Match 9.5%; Score 73.5; DB 1; Length 433;
 Best Local Similarity 22.7%; Pred. No. 18;
 Matches 35; Conservative 17; Mismatches 45; Indels 57; Gaps 6;

QY 14 VSGSALAGVVPWQCGGNGHNGGNS-----GPDSTLSIYQGSANAALALQSD 62
 DB 236 INGSGGGVQGMNATTNPSSGGSQIWDPNLQSRGPD 277
 QY 63 ARKYDQLVTVRYVTHEMAHAGGADNSTIELTONGFPNNATTIDOMAKNS-----DITV 115
 DB 278 -----SRV---EALYVGKALSS-----FTINNFDHMETCTYNKGVKEIK 318
 QY 116 GQYGGNNAALVNGTASDSSVMVROYFGNNAATAN 149
 DB 319 DDYKSTSGVQVPTKESN--NRPTFGSEKGSN 350

RESULT 7
 US-10-045-674A-527
 Sequence 527, Application US/10045674A

GENERAL INFORMATION:
 APPLICANT: LADNER, ROBERT C.
 APPLICANT: COHEN, EDWARD H.
 APPLICANT: NASTRI, HORACIO G.
 APPLICANT: ROONEY, KRISTIN L.
 APPLICANT: HOET, RENE
 APPLICANT: HOOGENDOORN, HENDRICUS R. J. M.
 TITLE OF INVENTION: NOVEL METHODS OF CONSTRUCTING LIBRARIES COMPRISING
 TITLE OF INVENTION: DISPLAYED AND/OR EXPRESSED MEMBERS OF A DIVERSE FAMILY
 TITLE OF INVENTION: OF PEPTIDES, POLYPEPTIDES OR PROTEINS AND THE NOVEL
 FILE REFERENCE: DYAX/002 CIP2
 CURRENT APPLICATION NUMBER: US/10/045,674A
 CURRENT FILING DATE: 2001-10-25
 PRIOR APPLICATION NUMBER: 06/198,069
 PRIOR FILING DATE: 2000-04-17
 PRIOR APPLICATION NUMBER: 09/837,306
 PRIOR FILING DATE: 2001-04-17
 NUMBER OF SEQ ID NOS: 635
 SOFTWARE: Patentin Ver. 2.1
 SEQ ID NO 527
 LENGTH: 533
 TYPE: PRT
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: Description of Artificial Sequence: Vector pC85
 US-10-045-674A-527

Query Match 9.5%; Score 73.5; DB 6; Length 533;
 Best Local Similarity 25.7%; Pred. No. 24;
 Matches 37; Conservative 15; Mismatches 47; Indels 45; Gaps 8;

QY 15 SCSALAGVVPWQCG--GGNHNCGNSGPDSTLSIYQ-YGSANAAL-----ALQSDA 63
 DB 352 SGGSGEGGSGEGGSGEGGSGEGGSGGSDPDEKMANNKAMTENADENALQSDA 411
 QY 64 R-KYDQLVTVRYVTHEMAHAGGADNSTIELTONGFPNNATTIDOMAKNSDITVGYQGN 122
 DB 412 KKKLDSVAT-----DYGAID-----GFIGDVS-----GLANGGATGDPAGSN 450
 QY 123 AALVNGTASDSSVMVROYFGNNA 146
 DB 451 SQMA-----QVGDGDS 462

RESULT 8
 PCT-US04-05654-590

Sequence 590, Application PC/TUS0405654
 GENERAL INFORMATION:
 APPLICANT: Sherman, Bradley K
 APPLICANT: Riechmann, Jose Luis
 APPLICANT: Jiang, Cai-Zhong
 APPLICANT: Haake, Jacqueline E
 APPLICANT: Haake, Volker
 APPLICANT: Creelman, Robert A
 APPLICANT: Ratcliffe, Oliver

```

APPLICANT: Adam, Luc J
APPLICANT: Reuber, T. Lynne
APPLICANT: Keddle, James
APPLICANT: Dubell III, Arnold N
APPLICANT: Pineda, Omaria
APPLICANT: Repetti, Peter
APPLICANT: Guterson, Neal
APPLICANT: Yu, Guo-liang
APPLICANT: Brown, Pierre E
APPLICANT: Kumoto, Roderick W
APPLICANT: Pilgrim, Maisha L
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES IN PLANTS
FILE REFERENCE: MBI-0047 PCT
CURRENT APPLICATION NUMBER: PCT/US04/05654
PRIOR FILING DATE: 2004-03-03
PRIOR APPLICATION NUMBER: 10/374,780
PRIOR FILING DATE: 2003-02-25
PRIOR APPLICATION NUMBER: 10/675,852
PRIOR FILING DATE: 2003-09-30
NUMBER OF SEQ ID NOS: 2950
SOFTWARE: PatentIn version 3.2
SEQ ID NO: 590
LENGTH: 321
TYPE: PRT
ORGANISM: Oryza sativa
FEATURE:
OTHER INFORMATION: G3384 Orthologous to G256
PCT-US04-05654-590

```

```

Query Match          9.4%; Score 73; DB 1; Length 321;
Best Local Similarity 23.5%; Pred. No. 14;
Matches 28; Conservative 19; Mismatches 42; Indels 30; Gaps 5;

```

```

Qy 10 AAIIVSSALAGVYQW---GGGHNHNGGSSGPDSTLIYQYGSANALALQSDARKY 66
Db 199 AAYVSSADNIALILQGMWPGGG---GGGNGKGPBEAS-----GSTSTATTQOCPQ-- 246
Qy 67 DQVTVRVVTHMAAGAGDNSTIELTONGFRNNATID--OMNAKNSDITVQYGSANNA 123
Db 247 -----CSGEGASASASASGSAALAAATQTPCSTIETSKMATGGAGGPA 292

```

```

RESULT 9
US-10-045-674A-594
Sequence 594, Application US/10045674A
GENERAL INFORMATION:
APPLICANT: LADNER, ROBERT C.
APPLICANT: COHEN, EDWARD H.
APPLICANT: NASTRI, HORACIO G.
APPLICANT: ROOKEY, KRISTIN L.
APPLICANT: HOEY, RENE
APPLICANT: HOOGENDOORN, HENDRICUS R. J. M.
TITLE OF INVENTION: NOVEL METHODS OF CONSTRUCTING LIBRARIES COMPRISING
TITLE OF INVENTION: DISPLAYED AND/OR EXPRESSED MEMBERS OF A DIVERSE FAMILY
TITLE OF INVENTION: OF PEPTIDES, POLYPEPTIDES OR PROTEINS AND THE NOVEL
FILE REFERENCE: DAX/002 CIP2
CURRENT APPLICATION NUMBER: US/10/045,674A
PRIOR FILING DATE: 2001-10-25
PRIOR APPLICATION NUMBER: 06/198,069
PRIOR FILING DATE: 2000-04-17
PRIOR APPLICATION NUMBER: 09/837,306
PRIOR FILING DATE: 2001-04-17
NUMBER OF SEQ ID NOS: 635
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO: 594
LENGTH: 434
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: M13-III
OTHER INFORMATION: protein sequence

```

US-10-045-674A-594

```

Query Match          9.4%; Score 72.5; DB 6; Length 434;
Best Local Similarity 25.6%; Pred. No. 23;
Matches 34; Conservative 17; Mismatches 43; Indels 39; Gaps 7;

```

```

Qy 27 GGGHNHNGGSSGPDSTLIYQYGSANALALQSDARKYDQVTVRVVTHMAAGAGAD 86
Db 257 GGGGSGSGSGSGEGG-----GSEGGSGSGSGSGDFD-----EKANANKGA- 298
Qy 87 NSTIELTONGFRNNATIDOMNAKNS-DITVQYGS-----GNALVYQTA-----SDS 133
Db 299 -----MTENDENAL---OSDAKGLDVAITDYGAIDGFIQVSGSLANGATGDFAGS 350
Qy 134 SYMYROYFGNNA 146
Db 351 NSQMAQVGDGDNQ 363

```

RESULT 10

```

US-10-417-884A-6047
Sequence 6047, Application US/10417884A
GENERAL INFORMATION:

```

```

APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
CORRESPONDENCE ADDRESS:
ADDRESS: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII

```

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/417,884A
FILING DATE: 17-Apr-2003
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532
FILING DATE: 30-Jun-1998
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997

```

```

ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 6047:
SEQUENCE CHARACTERISTICS:
LENGTH: 708 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...708
SEQUENCE DESCRIPTION: SEQ ID NO: 6047:
US-10-417-884A-6047

```

```

Query Match          9.3%; Score 72; DB 6; Length 708;
Best Local Similarity 24.9%; Pred. No. 48;

```

```

; SOFTWARE: FaetsBQ for Windows Version 4.0
; SEQ ID NO: 63
; LENGTH: 956
; TYPE: PRT
; ORGANISM: Bankia gouldii
US-10-093-037A-63

Query Match          9.2%; Score 71.5; DB 6; Length 956;
Best Local Similarity 19.8%; Pred. No. 79;
Matches 34; Conservative 22; Mismatches 67; Indels 49; Gaps 7;

Cy 12 IVSSGSAAGVPPQMGGGGNNHGGSSGSDSTLSIYOGSANNALALQSDARKYQDLVT 71
Db 318 LTPAGEIYKNIQNN-----DTETSTGPKTT-----QCTIECTRAEMETAQAGDIIIT 366
Cy 72 RVVTHEMAGAGADNSTIEL--TONGFRNNATI----- 103
Db 367 APGNVPQDKIQGAPNRSVLYLGSANGSTNPILIRGSATNPVPVSGLDYNNGYLLSTE 426
Cy 104 -DQNNAKSDITVGYG-----GNNAALVNOTASD--SSVWROYGFGNNA 146
Db 427 GDYNNIKIDFEKTKSGKIVLDNSNSKLNVLVHDIAGEAHLIRD-GSSNNS 477

RESULT 13
US-09-126-816C-6
; Sequence 6, Application US/09126816C
; GENERAL INFORMATION:
; APPLICANT: VON EICHEL-STREIBER, CHRISTOPH
; APPLICANT: BOQUET, PATRICE
; APPLICANT: THEBESTAM, MONICA
; TITLE OF INVENTION: METHOD OF INACTIVATION OF RAS SUBFAMILY PROTEINS AND
; FILE OF INVENTION: AGENTS THEREFOR
; FILE REFERENCE: 98501/254992/bec
; CURRENT APPLICATION NUMBER: US/09/126, 816C
; PRIOR FILING DATE: 1998-07-31
; PRIOR APPLICATION NUMBER: PCT/EP97/00426
; PRIOR FILING DATE: 1997-01-31
; PRIOR APPLICATION NUMBER: 96 101 469.3
; PRIOR FILING DATE: 1996-02-02
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO: 6
; LENGTH: 2364
; TYPE: PRT
; ORGANISM: Clostridium sordellii
US-09-126-816C-6

Query Match          9.2%; Score 71.5; DB 5; Length 2364;
Best Local Similarity 28.1%; Pred. No. 2, 5e+02;
Matches 27; Conservative 16; Mismatches 30; Indels 23; Gaps 4;

Cy 42 DSTLSIYOGSANNALALQSDARKYQDLVTRVYTHEMAGAGADNSTIELTON----- 95
Db 57 DNYNTYTKYKSGRNKAL-----KKEKELTMEVL-----ELKNSLTPEVKNIHFIMI 103
Cy 96 GFRNNAT-----IDQNNAKSDITVGYGYNNAALVN 127
Db 104 GGQINDRAINYINQMKVNSDYTVKVFDSNAFLIN 139

RESULT 14
US-10-767-701-37968
; Sequence 37968, Application US/10767701
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29

```

NUMBER OF SEQ ID NOS: 63128
SEQ ID NO 37968
LENGTH: 363
TYPE: PRT
ORGANISM: Sorghum bicolor
FEATURE:
OTHER INFORMATION: Clone ID: SORBI-28WAY03-C4212_1.p
US-10-767-701-37968

Query Match 9.2%; Score 71; DB 6; Length 363;
Best Local Similarity 24.1%; Pred. No. 25;
Matches 26; Conservative 15; Mismatches 49; Indels 18; Gaps 3;

QY 1 MKLKKVAFAIIVSGSALAGVPMWGGGNGNNGNSGSPSTLSIYQY-----SA 53
DB 29 LSLRRCACFCSCARS-----QOSNGANGNGTNGNGANGAPLVVDFDKTIVDCYSD 80

QY 54 NAALALQDARKYDQVTVTVTHMAHAGG--ADNSTIELTQNGFR 98
DB 81 NMVVDALGATRRFDDLLRHLPMNHAIDAMWGLHSDGKTAEPTLGS LR 128

RESULT 15
PCT-US04-05654-2748

Sequence 2748, Application PC/TUS0405654
GENERAL INFORMATION:

APPLICANT: Sherman, Bradley K
APPLICANT: Riechmann, Jose Luis
APPLICANT: Jiang, Cai Zhong
APPLICANT: Heard, Jacqueline E
APPLICANT: Haake, Volker
APPLICANT: Creelman, Robert A
APPLICANT: Ratcliffe, Oliver
APPLICANT: Adam, Luc J
APPLICANT: Reuber, T. Lynne
APPLICANT: Keddie, James
APPLICANT: Dubell III, Arnold N
APPLICANT: Pineda, Omaira
APPLICANT: Repetti, Peter
APPLICANT: Century, Karen
APPLICANT: Guttersen, Neal
APPLICANT: Yu, Guo-Hang
APPLICANT: Broun, Pierre E
APPLICANT: Kumimoto, Roderick W
APPLICANT: Pilgrim, Marsha L
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES IN PLANTS
FILE REFERENCE: MBI-0047 PCT
CURRENT APPLICATION NUMBER: PCT/US04/05654
CURRENT FILING DATE: 2004-03-03
PRIOR APPLICATION NUMBER: 10/374,780
PRIOR FILING DATE: 2003-02-25
PRIOR APPLICATION NUMBER: 10/675,852
PRIOR FILING DATE: 2003-09-30
NUMBER OF SEQ ID NOS: 2950
SOFTWARE: PatentIn version 3.2
SEQ ID NO 2748
LENGTH: 399
TYPE: PRT
ORGANISM: Arabidopsis thaliana
FEATURE:
OTHER INFORMATION: G1882
PCT-US04-05654-2748

Query Match 9.2%; Score 71; DB 1; Length 399;
Best Local Similarity 22.0%; Pred. No. 28;
Matches 44; Conservative 13; Mismatches 53; Indels 90; Gaps 8;

QY 27 GGGGNHNGGNSGSPDSTLSIYQY-----GSANAALALQDARKYD----- 67

DB 155 GGGGGSTSSGNSKQDTSATNDQYHHRAMANNQWGPSSSSSLSSYNAGLIPGHDH 214

QY 68 -----QVTVTVTHMAHAGGADNSTIELTQNGFR 98

DB 215 NSNNNNILGLSSLPLKLMPLDFTDNFTLQYGAVSAPSYHIGGSSGGAALL-NGF- 272
QY 99 NNATIDQW-----NAKNSDIT-VGQYGNNA--ALYNQTA 130
DB 273 -----DQMRPPATNQLPLGLDPEFDOHQWEOQNPQYGLVTSGGYRPKNIFHNLSSSS 327
QY 131 SDSVWVRQYGFQNNATYANQ 150
DB 328 SASSAMV-----TATASQ 340

Search completed: March 11, 2004, 19:14:17
Job time : 5.3 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:33:51 ; Search time 171.3 Seconds
(without alignments)
860.386 Million cell updates/sec

Title: US-09-543-407-22

Perfect score: 776
Sequence: 1 MLKLKVAFAIIVSGSALA.....DSSVMVRYQFGNNATANY 151Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 6019581 seqs, 976053577 residues

Total number of hits satisfying chosen parameters: 6019581

Minimum DB seq length: 0
Maximum DB seq length: 2000000000Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Pending Patents_AA_Main:*

1: /cgn2_6/ptodata/2/paa/PCTUS_COMB.pep:*
2: /cgn2_6/ptodata/2/paa/US06_COMB.pep:*
3: /cgn2_6/ptodata/2/paa/US07_COMB.pep:*
4: /cgn2_6/ptodata/2/paa/US080_COMB.pep:*
5: /cgn2_6/ptodata/2/paa/US081_COMB.pep:*
6: /cgn2_6/ptodata/2/paa/US082_COMB.pep:*
7: /cgn2_6/ptodata/2/paa/US083_COMB.pep:*
8: /cgn2_6/ptodata/2/paa/US084_COMB.pep:*
9: /cgn2_6/ptodata/2/paa/US085_COMB.pep:*
10: /cgn2_6/ptodata/2/paa/US086_COMB.pep:*
11: /cgn2_6/ptodata/2/paa/US087_COMB.pep:*
12: /cgn2_6/ptodata/2/paa/US088_COMB.pep:*
13: /cgn2_6/ptodata/2/paa/US089_COMB.pep:*
14: /cgn2_6/ptodata/2/paa/US090_COMB.pep:*
15: /cgn2_6/ptodata/2/paa/US091_COMB.pep:*
16: /cgn2_6/ptodata/2/paa/US092_COMB.pep:*
17: /cgn2_6/ptodata/2/paa/US093_COMB.pep:*
18: /cgn2_6/ptodata/2/paa/US094_COMB.pep:*
19: /cgn2_6/ptodata/2/paa/US095_COMB.pep:*
20: /cgn2_6/ptodata/2/paa/US096_COMB.pep:*
21: /cgn2_6/ptodata/2/paa/US097A_COMB.pep:*
22: /cgn2_6/ptodata/2/paa/US097B_COMB.pep:*
23: /cgn2_6/ptodata/2/paa/US098_COMB.pep:*
24: /cgn2_6/ptodata/2/paa/US099A_COMB.pep:*
25: /cgn2_6/ptodata/2/paa/US099B_COMB.pep:*
26: /cgn2_6/ptodata/2/paa/US100_COMB.pep:*
27: /cgn2_6/ptodata/2/paa/US101_COMB.pep:*
28: /cgn2_6/ptodata/2/paa/US102_COMB.pep:*
29: /cgn2_6/ptodata/2/paa/US103_COMB.pep:*
30: /cgn2_6/ptodata/2/paa/US104_COMB.pep:*
31: /cgn2_6/ptodata/2/paa/US106_COMB.pep:*
32: /cgn2_6/ptodata/2/paa/US107_COMB.pep:*
33: /cgn2_6/ptodata/2/paa/US107_COMB.pep:*
33: /cgn2_6/ptodata/2/paa/US107_COMB.pep:*

Prod. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
------------	-------	-------------	--------	----	-------------

1	776	100.0	151	19	US-09-543-407-22	Sequence 22, Appl
2	712	91.8	151	19	US-09-543-407-28	Sequence 28, Appl
3	682	87.9	151	19	US-09-543-407-5	Sequence 5, Appl
4	677	87.2	151	6	US-08-233-642A-57	Sequence 57, Appl
5	657	84.7	151	19	US-09-543-407-30	Sequence 30, Appl
6	619	79.8	151	19	US-09-543-407-20	Sequence 20, Appl
7	609	78.5	151	19	US-09-543-407-24	Sequence 24, Appl
8	604	77.8	151	19	US-09-543-407-12	Sequence 12, Appl
9	602	77.6	151	19	US-09-543-407-14	Sequence 14, Appl
10	601	77.4	151	19	US-09-543-407-26	Sequence 26, Appl
11	600	77.3	151	19	US-09-543-407-18	Sequence 18, Appl
12	595	76.7	131	19	US-09-543-407-18	Sequence 31, Appl
13	587	73.1	151	19	US-09-543-407-16	Sequence 16, Appl
14	521	67.1	151	19	US-09-543-407-7	Sequence 7, Appl
15	518	66.8	151	13	US-08-978-878-4	Sequence 4, Appl
16	518	66.8	151	21	US-09-741-873B-4	Sequence 4, Appl
17	516	66.5	151	33	US-60-352-946-2	Sequence 2, Appl
18	516	66.5	151	33	US-60-444-371-2	Sequence 2, Appl
19	497	64.0	120	6	US-08-233-642A-55	Sequence 55, Appl
20	465	59.9	158	16	US-09-252-691-5834	Sequence 5834, Ap
21	465	59.9	158	16	US-09-252-691C-5834	Sequence 5834, Ap
22	465	59.9	158	30	US-10-417-886-5834	Sequence 5834, Ap
23	463	59.7	109	19	US-09-543-407-34	Sequence 34, Appl
24	440	56.7	131	13	US-08-978-878-2	Sequence 2, Appl
25	440	56.7	131	21	US-09-741-873B-2	Sequence 2, Appl
26	336	43.3	109	19	US-09-543-407-35	Sequence 35, Appl
27	256	33.0	68	19	US-09-543-407-37	Sequence 37, Appl
28	215	27.7	48	19	US-09-543-407-39	Sequence 39, Appl
29	159.5	20.6	70	19	US-09-543-407-32	Sequence 32, Appl
30	104.5	13.5	186	16	US-09-252-691-5833	Sequence 5833, Ap
31	104.5	13.5	186	16	US-09-252-691C-5833	Sequence 5833, Ap
32	104.5	13.5	186	30	US-10-417-886-5833	Sequence 5833, Ap
33	97.5	12.6	520	1	PCT-US02-18256-21	Sequence 21, Appl
34	96	12.4	2308	33	US-60-161-992-68	Sequence 68, Appl
35	96	12.4	2309	20	US-09-614-150A-25488	Sequence 25488, A
36	96	12.4	2309	20	US-09-614-150A-25488	Sequence 25488, A
37	96	12.4	2309	33	US-60-191-637-25607	Sequence 25607, A
38	96	12.4	2309	33	US-60-191-637-25607	Sequence 25607, A
39	95.5	12.3	252	1	PCT-US03-13414-24	Sequence 24, Appl
40	95	12.2	445	29	US-10-369-493-20638	Sequence 20638, A
41	95	12.2	445	33	US-60-360-039-20638	Sequence 20638, A
42	95	12.2	597	1	PCT-US01-05992-146	Sequence 146, App
43	95	12.2	597	22	US-09-793-306-14	Sequence 146, App
44	94	12.1	271	30	US-10-437-963-147343	Sequence 147343,
45	93.5	12.0	145	21	US-09-739-449-8854	Sequence 8854, Ap

ALIGNMENTS

RESULT 1
US-09-543-407-22
Sequence 22, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543,407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 22
LENGTH: 151
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant salmonella enteritidis 3b atga
sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from GP63 of Yersinia major.

US-09-543-407-22

Query Match 100.0%; Score 776; DB 19; Length 151;
 Best Local Similarity 100.0%; Pred. No. 1.2e-74;
 Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLLKVAAPFAIVVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYQGSANALALQ 60
 DB 1 MLLKVAAPFAIVVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYQGSANALALQ 60
 QY 61 SDARKSETTTTOSGYNGADVQGADNYDQVTRVYTHEMAHADOMNAKNSDITVGYG 120
 DB 61 SDARKSETTTTOSGYNGADVQGADNYDQVTRVYTHEMAHADOMNAKNSDITVGYG 120
 QY 121 NNAALVNOTASDSVWVRQVGFNNATANQY 151
 DB 121 NNAALVNOTASDSVWVRQVGFNNATANQY 151

RESULT 2

US-09-543-407-28

Sequence 28, Application US/09543407

GENERAL INFORMATION:

APPLICANT: White, Aaron P.

APPLICANT: Doran, James L.

APPLICANT: Kay, William W.

TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR

FILE REFERENCE: 920043.406

CURRENT APPLICATION NUMBER: US/09/543.407

NUMBER OF SEQ ID NOS: 59

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 28

LENGTH: 151

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga

OTHER INFORMATION: sequence containing the replacement fragment

US-09-543-407-28

Query Match 91.8%; Score 712; DB 19; Length 151;
 Best Local Similarity 91.1%; Pred. No. 9.1e-68;
 Matches 144; Conservative 0; Mismatches 0; Indels 14; Gaps 2;

QY 1 MLLKVAAPFAIVVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYQGSANALALQ 60
 DB 1 MLLKVAAPFAIVVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYQGSANALALQ 60
 QY 61 SDARKSETTTTOSGYNGADVQGADNYDQVTRVYTHEMAHADOMNAKNSDI 113
 DB 61 SDARKSETTTTOSGYNGADVQGADNYDQVTRVYTHEMAHADOMNAKNSDI 113
 QY 114 TVGYGGNNALVNOTASDSVWVRQVGFNNATANQY 151
 DB 114 TVGYGGNNALVNOTASDSVWVRQVGFNNATANQY 151

RESULT 3

US-09-543-407-5

Sequence 5, Application US/09543407

GENERAL INFORMATION:

APPLICANT: White, Aaron P.

APPLICANT: Doran, James L.

APPLICANT: Kay, William W.

TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR

FILE REFERENCE: 920043.406

CURRENT APPLICATION NUMBER: US/09/543.407

Query Match 87.9%; Score 682; DB 19; Length 151;
 Best Local Similarity 90.1%; Pred. No. 1.5e-64;
 Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

CURRENT FILING DATE: 2000-04-05
 NUMBER OF SEQ ID NOS: 59
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 5
 LENGTH: 151
 TYPE: PRT
 ORGANISM: Salmonella enteritidis
 US-09-543-407-5

QY 1 MLLKVAAPFAIVVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYQGSANALALQ 60
 DB 1 MLLKVAAPFAIVVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYQGSANALALQ 60
 QY 61 SDARKSETTTTOSGYNGADVQGADNYDQVTRVYTHEMAHADOMNAKNSDITVGYG 120
 DB 61 SDARKSETTTTOSGYNGADVQGADNYDQVTRVYTHEMAHADOMNAKNSDITVGYG 120
 QY 121 NNAALVNOTASDSVWVRQVGFNNATANQY 151
 DB 121 NNAALVNOTASDSVWVRQVGFNNATANQY 151

RESULT 4

US-08-233-642A-57

Sequence 57, Application US/08233642A

GENERAL INFORMATION:

APPLICANT: Kay, William W.

APPLICANT: Clouthier, Sharon C.

APPLICANT: Doran, James L.

TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-

NUMBER OF SEQUENCES: 58

CORRESPONDENCE ADDRESSES:

ADDRESS: Seed and Berry

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: U.S.A.

ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/233.642A

FILING DATE: 26-APR-1994

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: King, Joshua

REGISTRATION NUMBER: 35,570

REFERENCE/DOCKET NUMBER: 920043.403C3

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

TELEX: 3723836 SEEDANBERRY

INFORMATION FOR SEQ ID NO: 57:

SEQUENCE CHARACTERISTICS:

LENGTH: 151 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-233-642A-57

Query Match 87.2%; Score 677; DB 6; Length 151;
 Best Local Similarity 90.1%; Pred. No. 5.2e-64;
 Matches 136; Conservative 2; Mismatches 13; Indels 0; Gaps 0;

```

QY      1 MKLLKVAAPFAIYVSSSALAGVPPQWGGGNNHNGGNSGPDSTLSIYOGSANAALALQ 60
      1 MKLLKVAAPFAIYVSSSALAGVPPQWGGGNNHNGGNSGPDSTLSIYOGSANAALALQ 60
DB      61 SDARKSETTITOSGYNGADVGGADNYDQLVTRVYVTHEMAHADQNNAKNSDITVGOYGG 120
      61 SDARKSETTITOSGYNGADVGGADNSTIELTONGFRNNATIDQNNAKNSDITVGOYGG 120
QY      121 NNAALVNQTAQSDSSVWVRQVGFNNATANOY 151
      121 NNPALVNQTAQSDSSVWVRQVGFNNATANOY 151
DB      121 NNPALVNQTAQSDSSVWVRQVGFNNATANOY 151

RESULT 5
US-09-543-407-30
; Sequence 30, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-30

Query Match      84.7%; Score 657; DB 19; Length 151;
Best Local Similarity 81.9%; Pred. No. 7.4e-62;
Matches 136; Conservative 0; Mismatches 0; Indels 30; Gaps 2;

QY      1 MKLLKVAAPFAIYVSSSALAGVPPQWGGGNNHNGGNSGPDSTLSIYOGSANAALALQ 60
      1 MKLLKVAAPFAIYVSSSALAGVPPQWGGGNNHNGGNSGPDSTLSIYOGSANAALALQ 60
DB      61 SDARKSETTITOSGYNGADVGGADNYDQLVTRVYVTHEMAHADQNNAKNSDITVGOYGG 105
      61 SDARKSETTITOSGYNGADVGGADNSTIELTONGFRNNATIDQNNAKNSDITVGOYGG 118
QY      106 WNAKNSDITVGOYGGNNAALVNQTAQSDSSVWVRQVGFNNATANOY 151
      119 -----GNNALVNQTAQSDSSVWVRQVGFNNATANOY 151
DB      119 -----GNNALVNQTAQSDSSVWVRQVGFNNATANOY 151

RESULT 6
US-09-543-407-20
; Sequence 20, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 151

```

```

; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-20

Query Match      79.8%; Score 619; DB 19; Length 151;
Best Local Similarity 74.6%; Pred. No. 8.9e-58;
Matches 129; Conservative 0; Mismatches 0; Indels 44; Gaps 2;

QY      1 MKLLKVAAPFAIYVSSSALAGVPPQWGGGNNHNGGNSGPDSTLSIYOGSANAALALQ 60
      1 MKLLKVAAPFAIYVSSSALAGVPPQWGGGNNHNGGNSGPDSTLSIYOGSANAALALQ 60
DB      61 SDARKSETTITOSGYNGADVGGADNYDQLVTRVYVTHEMAHADQNNAKNSDITVGOYGG 103
      61 SDARK-----YDQLVTRVYVTHEMAHADQNNAKNSDITVGOYGG 98
QY      104 -----DQNNAKNSDITVGOYGGNNAALVNQTAQSDSSVWVRQVGFNNATANOY 151
      99 NNATIDQNNAKNSDITVGOYGGNNAALVNQTAQSDSSVWVRQVGFNNATANOY 151
DB      99 NNATIDQNNAKNSDITVGOYGGNNAALVNQTAQSDSSVWVRQVGFNNATANOY 151

RESULT 7
US-09-543-407-24
; Sequence 24, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-24

Query Match      78.5%; Score 609; DB 19; Length 151;
Best Local Similarity 73.6%; Pred. No. 1.1e-56;
Matches 128; Conservative 0; Mismatches 0; Indels 46; Gaps 2;

QY      1 MKLLKVAAPFAIYVSSSALAGVPPQWGGGNNHNGGNSGPDSTLSIYOGSANAALALQ 60
      1 MKLLKVAAPFAIYVSSSALAGVPPQWGGGNNHNGGNSGPDSTLSIYOGSANAALALQ 60
DB      61 SDARKSETTITOSGYNGADVGGADNSTIELTONGFRNNATIDQNNAKNSDITVGOYGG 97
      61 SDARKSETTITOSGYNGADVGGADNSTIELTONGFRNNATIDQNNAKNSDITVGOYGG 120
QY      98 HEMAHADQNNAKNSDITVGOYGGNNAALVNQTAQSDSSVWVRQVGFNNATANOY 151
      121 HEMAHA-----NQTASDSSVWVRQVGFNNATANOY 151
DB      121 HEMAHA-----NQTASDSSVWVRQVGFNNATANOY 151

RESULT 8
US-09-543-407-12
; Sequence 12, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.

```

```

; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-12

```

```

Query Match          77.8%; Score 604; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 3.6e-56;
Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;

```

```

QY 1 MKLKVAAFAAIVVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
DB 1 MKLKVAAFAAIVVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
QY 61 SDARKSETTITGSGYGNGADVGGADNYDQVTRVVTHEMAADQWNAKNSDITVGYGG 120
DB 61 SDARKSETTITGSGYGNGADVGGADNSTIELTQNGFNNATIDQWNAKNSDITVGYGG 120
QY 121 NNAALVNOTASDSSVMRVQVGFNNATANY 151
DB 121 NNAALVNYDQVTRVVTHEMAHANNATANY 151

```

```

RESULT 9
US-09-543-407-14
; Sequence 14, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-14

```

```

Query Match          77.6%; Score 602; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 6e-56;
Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;

```

```

QY 1 MKLKVAAFAAIVVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
DB 1 MKLKVAAFAAIVVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
QY 61 SDARKSETTITGSGYGNGADVGGADNYDQVTRVVTHEMAADQWNAKNSDITVGYGG 120
DB 61 SDARKSETTITGSGYGNGADVGGADNSTIELTQNGFNNATIDQWNAKNSDITVGYGG 120

```

```

QY 121 NNAALVNOTASDSSVMRVQVGFNNATANY 151
DB 121 LVTRVVTHEMAHASVMRVQVGFNNATANY 151

```

```

RESULT 10
US-09-543-407-26
; Sequence 26, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-26

```

```

Query Match          77.4%; Score 601; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 7.7e-56;
Matches 123; Conservative 4; Mismatches 24; Indels 0; Gaps 0;

```

```

QY 1 MKLKVAAFAAIVVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
DB 1 MKLKVAAFAAIVVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
QY 61 SDARKSETTITGSGYGNGADVGGADNYDQVTRVVTHEMAADQWNAKNSDITVGYGG 120
DB 61 LVTRVVTHEMAHAGVNGADVGGADNSTIELTQNGFNNATIDQWNAKNSDITVGYGG 120
QY 121 NNAALVNOTASDSSVMRVQVGFNNATANY 151
DB 121 NNAALVNOTASDSSVMRVQVGFNNATANY 151

```

```

RESULT 11
US-09-543-407-18
; Sequence 18, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-18

```

```

Query Match          77.3%; Score 600; DB 19; Length 151;

```



```

; EARLIER APPLICATION NUMBER: US 07/347,189
; EARLIER FILING DATE: 1989-05-04
; EARLIER APPLICATION NUMBER: US 07/789,437
; EARLIER FILING DATE: 1991-11-06
; EARLIER APPLICATION NUMBER: US 07/970,846
; EARLIER FILING DATE: 1992-11-03
; EARLIER APPLICATION NUMBER: US 08/187,865
; EARLIER FILING DATE: 1994-01-28
; EARLIER APPLICATION NUMBER: US 08/318,519
; EARLIER FILING DATE: 1994-10-05
; EARLIER APPLICATION NUMBER: US 08/495,959
; EARLIER FILING DATE: 1995-06-28
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-08-978-878-4

```

```

Query Match      66.8%; Score 518; DB 13; Length 151;
Best Local Similarity 68.9%; Pred. No. 6,3e-47;
Matches 104; Conservative 18; Mismatches 29; Indels 0; Gaps 0;

```

```

QY      1 M K L K V A A F A I V S G S A L A G V P Q W G G G N H N G G N S G P D S T L I Y O G S A N A L A L Q 60
      1 M K L K V A I A I V F S G A V A G V P Q Y G G G N H G G G N S G P N S E L N I Y Q Y G G S A L A L Q 60
DB      61 S D A R S E T T I Q S G Y G N G A D V G G G A D N Y D Q L V T R V Y T H E M A H A D Q M A K N S D I T V G O Y G 120
      61 T D A R N S D L I T Q H G G G N A D V G G G S D S I D L T O R F G N S A T L D Q M N G K N S E M T V K O F G G 120
QY      121 N N A A L V N O T A S D S V M V R Q V G F G N N A T A N O Y 151
      121 G N G A A V D Q T A N S S V N V T Q V G F G N N A T A N O Y 151
DB

```

Search completed: March 11, 2004, 19:13:12
 Job time : 171.3 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:35:05 ; Search time 5.3 Seconds
(without alignments)
376.014 Million cell updates/sec

Title: US-09-543-407-22

Perfect score: 776
Sequence: 1 MLKLKVAFAIIVSGSALA.....DSSVMRYQVFGNNATANY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 125546 seqs, 13197846 residues

Total number of hits satisfying chosen parameters: 125546

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents_AA_New.*
1: /cgn2_6/prodata/1/paa/PCT_NEW_COMB.pep.*
2: /cgn2_6/prodata/1/paa/US06_NEW_COMB.pep.*
3: /cgn2_6/prodata/1/paa/US07_NEW_COMB.pep.*
4: /cgn2_6/prodata/1/paa/US08_NEW_COMB.pep.*
5: /cgn2_6/prodata/1/paa/US09_NEW_COMB.pep.*
6: /cgn2_6/prodata/1/paa/US10_NEW_COMB.pep.*
7: /cgn2_6/prodata/1/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	518	66.8	151	US-09-741-873C-4	Sequence 4, Appl1
2	440	56.7	131	US-09-741-873C-2	Sequence 2, Appl1
3	80.5	10.4	234	US-10-767-701-45603	Sequence 45603, A
4	80	10.3	430	US-10-451-677A-314	Sequence 314, App
5	79.5	10.2	245	US-10-779-461-19	Sequence 19, Appl
6	79	10.2	250	US-10-779-461-6	Sequence 6, Appl1
7	78.5	10.1	179	US-10-767-701-35342	Sequence 35342, A
8	78.5	10.1	251	US-10-767-701-32417	Sequence 32417, A
9	77.5	10.0	382	US-10-771-241-299	Sequence 299, App
10	77.5	10.0	1327	PCT-US04-02338-49	Sequence 49, Appl
11	77	9.9	250	US-10-779-461-40	Sequence 40, Appl
12	75.5	9.7	134	US-10-767-701-61422	Sequence 61422, A
13	75.5	9.7	424	US-10-045-674A-551	Sequence 527, App
14	75.5	9.7	533	US-10-045-674A-557	Sequence 527, App
15	75	9.7	174	US-10-767-701-56903	Sequence 56903, A
16	75	9.7	246	US-10-779-461-41	Sequence 41, Appl
17	75	9.6	399	PCT-US04-05654-2748	Sequence 2748, Ap
18	74.5	9.6	358	US-10-767-701-40444	Sequence 40444, A
19	73.5	9.5	351	PCT-US04-05654-1755	Sequence 1755, Ap
20	73	9.4	253	US-10-779-461-18	Sequence 18, Appl
21	73	9.4	1328	US-09-830-230A-631	Sequence 631, App
22	72.5	9.3	434	US-10-045-674A-554	Sequence 594, App
23	72.5	9.3	1871	PCT-US03-02038-26	Sequence 26, Appl
24	71.5	9.2	175	US-10-767-701-35046	Sequence 35046, A
25	71.5	9.2	179	US-10-767-701-42696	Sequence 42696, A
26	71.5	9.2	443	US-10-100-683-7608	Sequence 7608, Ap

27	71.5	9.2	673	US-10-417-888A-5134	Sequence 5134, Ap
28	71.5	9.2	1160	US-10-603-150-2	Sequence 2, Appl1
29	71	9.1	159	US-10-190-902B-10	Sequence 10, Appl1
30	71	9.1	250	US-10-779-461-22	Sequence 22, Appl
31	71	9.1	321	PCT-US04-05654-590	Sequence 590, App
32	71	9.1	523	US-10-641-678-45	Sequence 45, Appl
33	70.5	9.1	171	US-10-767-701-61575	Sequence 61575, A
34	70.5	9.1	346	PCT-US04-02242-58	Sequence 58, Appl
35	70.5	9.1	622	PCT-US04-05654-1743	Sequence 1743, Ap
36	70	9.0	245	US-10-779-461-59	Sequence 59, Appl
37	70	9.0	251	US-10-779-461-30	Sequence 30, Appl
38	70	9.0	485	PCT-US04-05654-1844	Sequence 1844, Ap
39	70	9.0	624	PCT-US04-05654-1466	Sequence 1466, Ap
40	70	9.0	1621	US-10-603-163-2	Sequence 2, Appl1
41	69.5	9.0	178	US-10-767-701-32700	Sequence 32700, A
42	69	8.9	182	US-10-767-701-44940	Sequence 44940, A
43	69	8.9	588	PCT-US04-02188-187	Sequence 187, App
44	69	8.9	588	US-10-764-425-187	Sequence 187, App
45	69	8.9	956	US-10-093-037A-63	Sequence 63, Appl

ALIGNMENTS

```
RESULT 1
US-09-741-873C-4
; Sequence 4, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibrinectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873C-4

Query Match
Best Local Similarity 66.8%; Score 518; DB 5; Length 151;
Matches 104; Conservative 18; Mismatches 29; Indels 0; Gaps 0;

QY 1 MLKLKVAFAIIVSGSALAGVVPQGGGNGGNSGSPSTLSIYYGGSNAAALAQ 60
  |||||
DB 1 MLKLKVAFAIIVSGSALAGVVPQGGGNGGNSGSPSTLSIYYGGSNAAALAQ 60
  |||||

QY 61 SPARKSETTITGSGNGADVGQAGADNYQLATRVVTHMMAHADQNAKNSDITVGQYGG 120
  |||||
DB 61 TPARKSDELITTHGGGNGADVGQAGADNYQLATRVVTHMMAHADQNAKNSDITVGQYGG 120
  |||||

QY 121 NNAALVNTASDSVVRQVFGNNATANY 151
  |||||
DB 121 GNGAALVDQASNSVVRQVFGNNATANY 151
  |||||
```

```

Query Match: 10.3%; Score 80; DB 6; Length 430;
Best Local Similarity 24.2%; Pred. No. 6.8;
Matches 30; Conservative 20; Mismatches 58; Indels 16; Gaps 5

QY      28 GCGNHNCGNSGSPSTLTIYQGS-ANMALALQSPARKSETTITQSGYGNAGDVGQGA- 85
Db      139 GSSNNNDYGSNNND-----YGSNNNDYGSNNNDSSYSSNNKKSSTGSSNNDSYGSN 192
          | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY      86 --DNVDLVTRVTHNAHADQWNAKNSDITVQYGGNNAALVNOQTASDS--SWVRQYQ 141
Db      193 NDDSYGSNNKKKSSYGSNNDSYGSNNDD---SYGSNNDSYGSNNNDSSYSSNNKKS 248
          : : |
          142 FGNN 145
          : : |
          249 YGSN 252

RESULT 5
US-10-779-461-19
; Sequence 19, Application US/10779461
; GENERAL INFORMATION:
; APPLICANT: Morton, Philip A
; TITLE OF INVENTION: ANTIBODIES TO C-MET FOR THE TREATMENT OF CANCERS
; FILE REFERENCE: 00980/1
; CURRENT APPLICATION NUMBER: US/10/779,461
; CURRENT FILING DATE: 2004-02-13
; PRIOR APPLICATION NUMBER: 60/447,073
; PRIOR FILING DATE: 2003-02-13

```


; SEQ ID NO 299
; LENGTH: 382
; TYPE: PRT
; ORGANISM: E. coli
US-10-771-241-299

Query Match 10.0%; Score 77.5; DB 6; Length 382;
Best Local Similarity 26.1%; Pred. No. 9.7;
Matches 30; Conservative 13; Mismatches 37; Indels 35; Gaps 5;

QY 30 GNHNGGSSGPDSTLSTIYQ--YGSANAALALOSDARKSETTITQSGYGNAGADVGGAD 86
DB 158 GATHGAGTSSGTSGYSSSNTRGAYNTVYGRSGDS-----SQIYGMSSGGI----- 205

QY 87 NYDQLVTRVTHEMAHADQWNAKNSDITVGYGYNNAALVNGTASDSSVWVRQY 141
DB 206 -----IHHAD-----GITFGQPLGDTMVLVAPGADNVKXENQTG 240

RESULT 10
PCT-US04-02338-49
; Sequence 49, Application PC/TUS0402338
; GENERAL INFORMATION:

; APPLICANT: EXELIXIS, INC.
; TITLE OF INVENTION: MAPCAXS AS MODIFIERS OF THE APC AND AXIN PATHWAYS AND METHODS OF
; FILE REFERENCE: EX04-003C-PC
; CURRENT APPLICATION NUMBER: PCT/US04/02338
; PRIOR FILING DATE: 2004-01-28
; PRIOR APPLICATION NUMBER: US60/443,484
; PRIOR FILING DATE: 2003-01-29
; PRIOR APPLICATION NUMBER: US60/447,358
; PRIOR FILING DATE: 2003-02-11
; PRIOR APPLICATION NUMBER: US60/461,789
; PRIOR FILING DATE: 2003-04-10
; PRIOR APPLICATION NUMBER: US60/470,684
; PRIOR FILING DATE: 2003-05-14
; PRIOR APPLICATION NUMBER: US60/479,650
; PRIOR FILING DATE: 2003-06-19
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 49
; LENGTH: 1327
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US04-02338-49

Query Match 10.0%; Score 77.5; DB 1; Length 1327;
Best Local Similarity 33.3%; Pred. No. 44;
Matches 27; Conservative 12; Mismatches 31; Indels 11; Gaps 3;

QY 6 VAAFAAI-VVSGSALAGVDPQWGGGNGHNGSGPDSITLSTIYGSANAALALOSDAR 64
DB 99 VAAAPVPAVSTSSAAGVAPNPAGSGSSSSSPTSS--SSSSPSSPSSSLAESPEA- 156

QY 65 KSETTITQSGYGNAGADVGGGA 85
DB 157 -----AGVSTAPPLGPGA 169

RESULT 11
US-10-779-461-40
; Sequence 40, Application US/10779461
; GENERAL INFORMATION:

; APPLICANT: MORTON, PHILIP A
; TITLE OF INVENTION: ANTIBODIES TO C-MET FOR THE TREATMENT OF CANCERS
; FILE REFERENCE: 00980/1
; CURRENT APPLICATION NUMBER: US/10/779,461
; PRIOR FILING DATE: 2004-02-13
; PRIOR APPLICATION NUMBER: 60/447,073
; PRIOR FILING DATE: 2003-02-13
; NUMBER OF SEQ ID NOS: 161
; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 40
; LENGTH: 250
; TYPE: PRT
; ORGANISM: artificial
; FEATURE:
; OTHER INFORMATION: phage display generated human antibody
US-10-779-461-40

Query Match 9.9%; Score 77; DB 6; Length 250;
Best Local Similarity 30.8%; Pred. No. 6.4;
Matches 28; Conservative 8; Mismatches 31; Indels 24; Gaps 4;

QY 9 FAIIVSGSALAGVDPQW-----GGGNHNGGSSGPDSTLSTIYGSANAALAL 59
DB 95 YCAISPLRGITLADVPFWGQGITLVYSSGGSGSGGSGGG-----GSAQSALIQ 144

QY 60 QSDARKSE-TTITQSGYGNAGADVGGADNYD 89
DB 145 PASAGSPGQSITRISCTGTSPDGR----YD 171

RESULT 12
US-10-767-701-61422
; Sequence 61422, Application US/10767701
; GENERAL INFORMATION:

; APPLICANT: KOVALIC, DAVID K.
; APPLICANT: ZHOU, YIHUA
; APPLICANT: CAO, YONGWEI
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 61422
; LENGTH: 134
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: 9306804.pep
US-10-767-701-61422

Query Match 9.7%; Score 75.5; DB 6; Length 134;
Best Local Similarity 23.4%; Pred. No. 4.1;
Matches 29; Conservative 22; Mismatches 38; Indels 35; Gaps 6;

QY 11 AIVVSGSAL-----AGVDPQWGGG-----NHNHNGGSSGPDSTLSTI 47
DB 16 AAVILGASLDTSPVTFHAGYMLFPGGNIIVPSPGRSVRLKIDRTTSGFVS-----KSA 70

QY 48 YOYGSANAALALOSDARKSETTITQSGYGNAGADVGGADNYDQLVTRVTHEMAHADQW 107
DB 71 YHHGFPSSIKLPD---DTAGVVAFLSNADVPPG--NHDEVDFELGNRRGH--EWR 123

QY 108 AANS 111
DB 124 VQTN 127

RESULT 13
US-10-045-674A-591
; Sequence 591, Application US/10045674A
; GENERAL INFORMATION:

; APPLICANT: LADNER, ROBERT C.
; APPLICANT: COHEN, EDWARD H.
; APPLICANT: NASTRI, HORACIO G.
; APPLICANT: ROONEY, KRISTIN L.
; APPLICANT: HOEFT, RENE
; APPLICANT: HOOGENDOORN, HENDRICUS R. J. M.
; TITLE OF INVENTION: NOVEL METHODS OF CONSTRUCTING LIBRARIES COMPRISING
; TITLE OF INVENTION: DISPLAYED AND/OR EXPRESSED MEMBERS OF A DIVERSE FAMILY
; TITLE OF INVENTION: OF PEPTIDES, POLYPEPTIDES OR PROTEINS AND THE NOVEL
LIBRARIES

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:33:51 ; Search time 171.3 Seconds
(without alignments)
860.386 Million cell updates/sec

Title: US-09-543-407-24

Perfect score: 775
Sequence: 1 MKLLKVAFAIVSGSALA.....DSSVMRVQGFNNATANY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 6019581 seqs, 976053577 residues

Total number of hits satisfying chosen parameters: 6019581

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending Patents_AA_Main:*

1: /cgn2_6/ptodata/2/paa/PTUTUS_COMB.pep.*
2: /cgn2_6/ptodata/2/paa/US06_COMB.pep.*
3: /cgn2_6/ptodata/2/paa/US07_COMB.pep.*
4: /cgn2_6/ptodata/2/paa/US080_COMB.pep.*
5: /cgn2_6/ptodata/2/paa/US081_COMB.pep.*
6: /cgn2_6/ptodata/2/paa/US082_COMB.pep.*
7: /cgn2_6/ptodata/2/paa/US083_COMB.pep.*
8: /cgn2_6/ptodata/2/paa/US084_COMB.pep.*
9: /cgn2_6/ptodata/2/paa/US085_COMB.pep.*
10: /cgn2_6/ptodata/2/paa/US086_COMB.pep.*
11: /cgn2_6/ptodata/2/paa/US087_COMB.pep.*
12: /cgn2_6/ptodata/2/paa/US088_COMB.pep.*
13: /cgn2_6/ptodata/2/paa/US089_COMB.pep.*
14: /cgn2_6/ptodata/2/paa/US090_COMB.pep.*
15: /cgn2_6/ptodata/2/paa/US091_COMB.pep.*
16: /cgn2_6/ptodata/2/paa/US092_COMB.pep.*
17: /cgn2_6/ptodata/2/paa/US093_COMB.pep.*
18: /cgn2_6/ptodata/2/paa/US094_COMB.pep.*
19: /cgn2_6/ptodata/2/paa/US095_COMB.pep.*
20: /cgn2_6/ptodata/2/paa/US096_COMB.pep.*
21: /cgn2_6/ptodata/2/paa/US097A_COMB.pep.*
22: /cgn2_6/ptodata/2/paa/US097B_COMB.pep.*
23: /cgn2_6/ptodata/2/paa/US098_COMB.pep.*
24: /cgn2_6/ptodata/2/paa/US099A_COMB.pep.*
25: /cgn2_6/ptodata/2/paa/US099B_COMB.pep.*
26: /cgn2_6/ptodata/2/paa/US100_COMB.pep.*
27: /cgn2_6/ptodata/2/paa/US101_COMB.pep.*
28: /cgn2_6/ptodata/2/paa/US102_COMB.pep.*
29: /cgn2_6/ptodata/2/paa/US103_COMB.pep.*
30: /cgn2_6/ptodata/2/paa/US104_COMB.pep.*
31: /cgn2_6/ptodata/2/paa/US106_COMB.pep.*
32: /cgn2_6/ptodata/2/paa/US107_COMB.pep.*
33: /cgn2_6/ptodata/2/paa/US60_COMB.pep.*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
------------	-------	--------------------	----	-------------

1	775	100.0	151 19	US-09-543-407-24	Sequence 24, Appl
2	714	92.1	151 19	US-09-543-407-14	Sequence 14, Appl
3	700	90.3	151 19	US-09-543-407-30	Sequence 30, Appl
4	690	89.0	151 19	US-09-543-407-5	Sequence 5, Appl
5	689	88.9	151 6	US-08-233-642A-57	Sequence 57, Appl
6	655	84.5	151 19	US-09-543-407-12	Sequence 12, Appl
7	609	78.6	151 19	US-09-543-407-22	Sequence 22, Appl
8	608	78.6	151 19	US-09-543-407-26	Sequence 26, Appl
9	606	78.2	151 19	US-09-543-407-18	Sequence 18, Appl
10	603	77.8	151 19	US-09-543-407-20	Sequence 20, Appl
11	603	77.8	151 19	US-09-543-407-31	Sequence 31, Appl
12	575	74.2	151 19	US-09-543-407-28	Sequence 28, Appl
13	575	74.2	151 19	US-09-543-407-16	Sequence 16, Appl
14	533	68.8	151 19	US-09-543-407-7	Sequence 7, Appl
15	530	68.4	151 13	US-08-978-878-4	Sequence 4, Appl
16	530	68.4	151 21	US-09-741-873B-4	Sequence 4, Appl
17	528	68.1	151 33	US-60-352-946-2	Sequence 2, Appl
18	528	68.1	151 33	US-60-444-371-2	Sequence 2, Appl
19	505	65.2	120 6	US-08-233-642A-55	Sequence 55, Appl
20	475	61.3	158 16	US-09-252-691-5834	Sequence 5834, Ap
21	475	61.3	158 16	US-09-252-691C-5834	Sequence 5834, Ap
22	475	61.3	158 30	US-10-417-886-5834	Sequence 5834, Ap
23	471	60.8	109 19	US-09-543-407-34	Sequence 34, Appl
24	452	58.3	131 13	US-08-978-878-2	Sequence 2, Appl
25	452	58.3	131 21	US-09-741-873B-2	Sequence 2, Appl
26	348	44.9	109 19	US-09-543-407-35	Sequence 35, Appl
27	347	44.8	68 19	US-09-543-407-37	Sequence 37, Appl
28	248.5	32.1	70 19	US-09-543-407-32	Sequence 32, Appl
29	237	30.6	48 19	US-09-543-407-39	Sequence 39, Appl
30	111	14.3	445 29	US-10-368-493-20638	Sequence 20638, A
31	111	14.3	445 33	US-60-360-039-20638	Sequence 20638, A
32	107	13.8	186 16	US-09-252-691-5833	Sequence 5833, Ap
33	107	13.8	186 16	US-09-252-691C-5833	Sequence 5833, Ap
34	107	13.8	186 30	US-10-417-886-5833	Sequence 5833, Ap
35	98.5	12.7	151 19	US-09-543-407-6	Sequence 6, Appl
36	98	12.6	903 24	US-09-914-543-44	Sequence 44, Appl
37	96.5	12.5	151 19	US-09-543-407-8	Sequence 8, Appl
38	92.5	11.9	850 20	US-09-614-150A-24084	Sequence 24084, A
39	92.5	11.9	850 20	US-09-614-150A-24084	Sequence 24084, A
40	92.5	11.9	850 33	US-60-191-637-24179	Sequence 24179, A
41	92.5	11.9	850 33	US-60-191-637-24179	Sequence 24179, A
42	92.5	11.9	1028 20	US-09-614-150-14916	Sequence 14916, A
43	92.5	11.9	1028 20	US-09-614-150A-14916	Sequence 14916, A
44	92.5	11.9	1028 33	US-60-167-217-14978	Sequence 14978, A
45	92.5	11.9	1028 33	US-60-173-464-12194	Sequence 12194, A

ALIGNMENTS

RESULT 1
US-09-543-407-24
Sequence 24, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIBRILL SYSTEM FOR
PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543,407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 24
LENGTH: 151
TYPE: PRF
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
OTHER INFORMATION: sequence containing the replacement fragment
encoding PT3 from Gp63 of Leishmania major.

US-09-543-407-24

Query Match 100.0%; Score 775; DB 19; Length 151;
Best Local Similarity 100.0%; Pred. No. 2,7e-74;
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLKVAAFAAIVVSGSALAGVVPOMGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
DB 1 MKLKVAAFAAIVVSGSALAGVVPOMGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
QY 61 SPARKSETTITQSGYGNAGADVQGGADNSTIELTONGFRNNATIDQWNAKNDQLVTRVVT 120
DB 61 SPARKSETTITQSGYGNAGADVQGGADNSTIELTONGFRNNATIDQWNAKNDQLVTRVVT 120
QY 121 HEMAHANOTASDSSVWVRQVGFNNATANOY 151
DB 121 HEMAHANOTASDSSVWVRQVGFNNATANOY 151

RESULT 2

US-09-543-407-14

; Sequence 14; Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collins, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FaSTSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of *Leishmania major*.
US-09-543-407-14

Query Match 92.1%; Score 714; DB 19; Length 151;
Best Local Similarity 91.1%; Pred. No. 9,2e-68;
Matches 144; Conservative 0; Mismatches 0; Indels 14; Gaps 2;

QY 1 MKLKVAAFAAIVVSGSALAGVVPOMGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
DB 1 MKLKVAAFAAIVVSGSALAGVVPOMGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
QY 61 SPARKSETTITQSGYGNAGADVQGGADNSTIELTONGFRNNATIDQWNAKNDQLVTRVVT 120
DB 61 SPARKSETTITQSGYGNAGADVQGGADNSTIELTONGFRNNATIDQWNAKNDQLVTRVVT 120
QY 114 LTRVVTHEMAHANOTASDSSVWVRQVGFNNATANOY 151
DB 121 LTRVVTHEMAHANOTASDSSVWVRQVGFNNATANOY 151

RESULT 3

US-09-543-407-30

; Sequence 30; Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collins, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407

CURRENT FILING DATE: 2000-04-05

NUMBER OF SEQ ID NOS: 59

SOFTWARE: FaSTSeq for Windows Version 4.0

SEQ ID NO 30

LENGTH: 151

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afgA

OTHER INFORMATION: sequence containing the replacement fragment

OTHER INFORMATION: encoding PT3 from GP63 of *Leishmania major*.
US-09-543-407-30

Query Match 90.3%; Score 700; DB 19; Length 151;
Best Local Similarity 89.9%; Pred. No. 2,9e-66;
Matches 143; Conservative 0; Mismatches 0; Indels 16; Gaps 2;

QY 1 MKLKVAAFAAIVVSGSALAGVVPOMGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
DB 1 MKLKVAAFAAIVVSGSALAGVVPOMGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
QY 61 SPARKSETTITQSGYGNAGADVQGGADNSTIELTONGFRNNATIDQWNAKNDQLVTRVVT 120
DB 61 SPARKSETTITQSGYGNAGADVQGGADNSTIELTONGFRNNATIDQWNAKNDQLVTRVVT 120
QY 121 HEMAHANOTASDSSVWVRQVGFNNATANOY 151
DB 113 HEMAHANOTASDSSVWVRQVGFNNATANOY 151

RESULT 4

US-09-543-407-5

; Sequence 5; Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collins, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FaSTSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 151
; TYPE: PRT
; ORGANISM: *Salmonella enteritidis*
US-09-543-407-5

Query Match 89.0%; Score 690; DB 19; Length 151;
Best Local Similarity 91.4%; Pred. No. 3,4e-65;
Matches 138; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

QY 1 MKLKVAAFAAIVVSGSALAGVVPOMGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
DB 1 MKLKVAAFAAIVVSGSALAGVVPOMGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
QY 61 SPARKSETTITQSGYGNAGADVQGGADNSTIELTONGFRNNATIDQWNAKNDQLVTRVVT 120
DB 61 SPARKSETTITQSGYGNAGADVQGGADNSTIELTONGFRNNATIDQWNAKNDQLVTRVVT 120
QY 121 HEMAHANOTASDSSVWVRQVGFNNATANOY 151
DB 121 HEMAHANOTASDSSVWVRQVGFNNATANOY 151

RESULT 5

US-08-233-642A-57

; Sequence 57; Application US/08233642A
; GENERAL INFORMATION:
; APPLICANT: Kay, William W.

APPLICANT: Collinson, S. Karen
APPLICANT: Cloutier, Sharon C.
APPLICANT: Doran, James L.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
TITLE OF INVENTION: BASED VACCINES
NUMBER OF SEQUENCES: 58
NUMBER OF SEQUENCES: -
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed and Berry
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: U.S.A.
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: PC floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233,642A
FILING DATE: 26-APR-1994
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: King, Joshua
REGISTRATION NUMBER: 35,570
REFERENCE/DOCKET NUMBER: 920043.403C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEEDBERRY
INFORMATION FOR SEQ ID NO: 57:
SEQUENCE CHARACTERISTICS:
LENGTH: 151 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-233-642A-57

Query Match 88.9%; Score 689; DB 6; Length 151;
Best Local Similarity 91.4%; Pred. No. 4.4e-65;
Matches 138; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

QY 1 MLLKVAAPFAIVSGSALAGVPPQGGGNNHGGSSGPDSTLSIYOGSANAALALQ 60
DB 1 MLLKVAAPFAIVSGSALAGVPPQGGGNNHGGSSGPDSTLSIYOGSANAALALQ 60
QY 61 SPARKSETTITSGYNGADVGQADNSTIELTONGFRNNATIDQNNAKNDITVQYGG 120
DB 61 SPARKSETTITSGYNGADVGQADNSTIELTONGFRNNATIDQNNAKNDITVQYGG 120
QY 121 HEMAHANOTASDSVWVROVGFNNATANY 151
DB 121 NNPAIVNOTASDSVWVROVGFNNATANY 151

RESULT 6
US-09-543-407-12
Sequence 12, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543,407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12
LENGTH: 151

TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-12

Query Match 84.5%; Score 655; DB 19; Length 151;
Best Local Similarity 79.8%; Pred. No. 1.9e-61;
Matches 134; Conservative 0; Mismatches 0; Indels 34; Gaps 2;

QY 1 MLLKVAAPFAIVSGSALAGVPPQGGGNNHGGSSGPDSTLSIYOGSANAALALQ 60
DB 1 MLLKVAAPFAIVSGSALAGVPPQGGGNNHGGSSGPDSTLSIYOGSANAALALQ 60
QY 61 SPARKSETTITSGYNGADVGQADNSTIELTONGFRNNATIDQNNAKNDITVQYGG 109
DB 61 SPARKSETTITSGYNGADVGQADNSTIELTONGFRNNATIDQNNAKNDITVQYGG 120
QY 110 -----NYDQLVTRVTHEMAHNOTASDSVWVROVGFNNATANY 151
DB 121 NNPAIVNOTASDSVWVROVGFNNATANY 151

RESULT 7
US-09-543-407-22
Sequence 22, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543,407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 22
LENGTH: 151
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-22

Query Match 78.6%; Score 609; DB 19; Length 151;
Best Local Similarity 73.6%; Pred. No. 1.6e-56;
Matches 128; Conservative 0; Mismatches 0; Indels 46; Gaps 2;

QY 1 MLLKVAAPFAIVSGSALAGVPPQGGGNNHGGSSGPDSTLSIYOGSANAALALQ 60
DB 1 MLLKVAAPFAIVSGSALAGVPPQGGGNNHGGSSGPDSTLSIYOGSANAALALQ 60
QY 61 SPARKSETTITSGYNGADVGQADNSTIELTONGFRNNATIDQNNAKNDITVQYGG 120
DB 61 SPARKSETTITSGYNGADVGQADNSTIELTONGFRNNATIDQNNAKNDITVQYGG 120
QY 121 HEMAHANOTASDSVWVROVGFNNATANY 151
DB 98 HEMAHANOTASDSVWVROVGFNNATANY 151

RESULT 8
US-09-543-407-26
Sequence 26, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.

```

; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-26

```

```

Query Match          78.6%; Score 609; DB 19; Length 151;
Best Local Similarity 82.1%; Pred. No. 1.6e-56;
Matches 124; Conservative 4; Mismatches 23; Indels 0; Gaps 0;

```

```

QY      1 MLLKVAAPAAIVVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
      1 MLLKVAAPAAIVVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALYDQ 60
DB      61 SDARKSETTTGSGYGADVGAGADNSTIELTONGFRNNATTIDQWNAKNDITVGYGG 120
      61 LVTRVVTHEMAHAGVGNADVGAGADNSTIELTONGFRNNATTIDQWNAKNSDITVGYGG 120
QY      121 HEMAHANQTASDSSVMVRQVGFNNATANQY 151
      121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151
DB

```

RESULT 9

```

US-09-543-407-18
; Sequence 18, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-18

```

```

Query Match          78.5%; Score 608; DB 19; Length 151;
Best Local Similarity 82.1%; Pred. No. 2.1e-56;
Matches 124; Conservative 5; Mismatches 22; Indels 0; Gaps 0;

```

```

QY      1 MLLKVAAPAAIVVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
      1 MLLKVAAPAAIVVSGSALAGVVPQWGGGNNHNGGNSGPDYDQLVTRVVTHEMAHALQ 60
DB      61 SDARKSETTTGSGYGADVGAGADNSTIELTONGFRNNATTIDQWNAKNDITVGYGG 120
      61 SDARKSETTTGSGYGADVGAGADNSTIELTONGFRNNATTIDQWNAKNSDITVGYGG 120
DB

```

```

QY      121 HEMAHANQTASDSSVMVRQVGFNNATANQY 151
      121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151
DB

```

RESULT 10

```

US-09-543-407-20
; Sequence 20, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-20

```

```

Query Match          78.2%; Score 606; DB 19; Length 151;
Best Local Similarity 82.1%; Pred. No. 3.4e-56;
Matches 124; Conservative 5; Mismatches 22; Indels 0; Gaps 0;

```

```

QY      1 MLLKVAAPAAIVVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
      1 MLLKVAAPAAIVVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
DB      61 SDARKSETTTGSGYGADVGAGADNSTIELTONGFRNNATTIDQWNAKNDITVGYGG 120
      61 SDARKDQLVTRVVTHEMAHAGVGNADVGAGADNSTIELTONGFRNNATTIDQWNAKNSDITVGYGG 120
QY      121 HEMAHANQTASDSSVMVRQVGFNNATANQY 151
      121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151
DB

```

RESULT 11

```

US-09-543-407-31
; Sequence 31, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Salmonella enteritidis
US-09-543-407-31

```

```

Query Match          77.8%; Score 603; DB 19; Length 131;
Best Local Similarity 90.1%; Pred. No. 5.9e-56;
Matches 118; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

```

```

QY      21 GVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQSDARKSETTTTGGYGAGAD 80

```

Db 1 GVPVQGGGNNHGGNSGPDSTLSIYQGSANALALQSDARKSETTTTGGYGNAD 60
QY 81 VGGAGNSTTELTONGFNNATIDQNNAKNYDQLVTRVVTHEMAHNOQTASDSVWVROV 140
Db 61 VGGAGNSTTELTONGFNNATIDQNNAKNSDITVQYGSNNAALVNOQTASDSVWVROV 120
QY 141 GFGNNATANOY 151
Db 121 GFGNNATANOY 131

RESULT 12
US-09-543-407-28
; Sequence 28, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
US-09-543-407-28

Query Match 77.4%; Score 600; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 1.5e-55;
Matches 123; Conservative 3; Mismatches 25; Indels 0; Gaps 0;
QY 1 MKLKVAAFAIIVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYQGSANALALQ 60
Db 1 MKLKVAAFAIIVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYQGSANALALQ 60
QY 61 SDARKSETTTTGGYGNADVGGAGNSTTELTONGFNNATIDQNNAKNYDQLVTRVVT 120
Db 61 SDARKSETTTTGGYGNADVGGAGNSTTELTONGFNNATIDQNNAKNSDITVQYGG 120
QY 121 HEMAHANQTSASDSVWVROVGFNNATANOY 151
Db 121 NNAALVNOQTASDSVWVROVGFNNATANOY 151

RESULT 13
US-09-543-407-16
; Sequence 16, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence

; FEATURE:
; OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of *Leishmania* major.
US-09-543-407-16

Query Match 74.2%; Score 575; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 7.2e-53;
Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;
QY 1 MKLKVAAFAIIVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYQGSANALALQ 60
Db 1 MKLKVAAFAIIVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYQGSANALALQ 60
QY 61 SDARKSETTTTGGYGNADVGGAGNSTTELTONGFNNATIDQNNAKNYDQLVTRVVT 120
Db 61 SDARKSETTTTGGYGNADVGGAGNSTTELTONGFNNATIDQNNAKNSDITVQYGG 120
QY 121 HEMAHANQTSASDSVWVROVGFNNATANOY 151
Db 121 NNAALVNOQTASDSVWVROVGFNNATANOY 151

RESULT 14
US-09-543-407-7
; Sequence 7, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 151
; TYPE: PRT
; ORGANISM: *Escherichia coli*
US-09-543-407-7

Query Match 68.8%; Score 533; DB 19; Length 151;
Best Local Similarity 70.2%; Pred. No. 2.3e-48;
Matches 106; Conservative 18; Mismatches 27; Indels 0; Gaps 0;
QY 1 MKLKVAAFAIIVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYQGSANALALQ 60
Db 1 MKLKVAAFAIIVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYQGSANALALQ 60
QY 61 SDARKSETTTTGGYGNADVGGAGNSTTELTONGFNNATIDQNNAKNYDQLVTRVVT 120
Db 61 TDARNSDLTTTGGYGNADVGGAGNSDITVQYGSNNAALVNOQTASDSVWVROV 120
QY 121 HEMAHANQTSASDSVWVROVGFNNATANOY 151
Db 121 GNGAANDQTSASDSVWVROVGFNNATANOY 151

RESULT 15
US-08-978-878-4
; Sequence 4, Application US/08978878
; GENERAL INFORMATION:
; APPLICANT: NORMAN, Staffan
; APPLICANT: OLSEN, Arne
; TITLE OF INVENTION: FIBRONECTIN BINDING PROTEIN AS WELL AS ITS PREPARATION
; FILE REFERENCE: 012889-081
; CURRENT APPLICATION NUMBER: US/08/978,878
; CURRENT FILING DATE: 1997-11-26
; EARLIER APPLICATION NUMBER: SE 8801723-1
; EARLIER FILING DATE: 1988-05-06

```

; EARLIER APPLICATION NUMBER: US 07/347,189
; EARLIER FILING DATE: 1989-05-04
; EARLIER APPLICATION NUMBER: US 07/789,437
; EARLIER FILING DATE: 1991-11-06
; EARLIER APPLICATION NUMBER: US 07/970,846
; EARLIER FILING DATE: 1992-11-03
; EARLIER APPLICATION NUMBER: US 08/187,865
; EARLIER FILING DATE: 1994-01-28
; EARLIER APPLICATION NUMBER: US 08/318,519
; EARLIER FILING DATE: 1994-10-05
; EARLIER APPLICATION NUMBER: US 08/495,959
; EARLIER FILING DATE: 1995-06-28
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-08-978-878-4

```

```

Query Match      68.4%; Score 530; DB 13; Length 151;
Best Local Similarity 69.5%; Pred. No. 4.8e-48;
Matches 105; Conservative 19; Mismatches 27; Indels 0; Gaps 0;

```

```

QY      1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNNHNGGNSGSPDSTLSIYQGSANALALQ 60
Db      1 MKLLKVAIAIAIVFSGSAVAGVVPQYGGGNGHGGGNNSGPNSSELNIYQGGNSALALQ 60
QY      61 SDARKSETTITQSGYGKADVDGQADNSTIELTONGFRNNATIDQWNAKNTDQLVTRVVT 120
Db      61 TDARNSDLTTTQHGGGNGADVDGQSDDSIDLTYRGFNGSATILDQWNGKNSGMTVKQFGG 120
QY      121 HEMAHANQTASDSVMTYRQVGFNNATATANOY 151
Db      121 GNGAAVDQTASNSSVNTYQVGFNNATATAHQY 151

```

Search completed: March 11, 2004, 19:13:13
 Job time : 172.3 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:35:05 ; Search time 5.3 Seconds
(without alignments)
376.014 Million cell updates/sec

Title: US-09-543-407-24

Perfect score: 775

Sequence: 1 MLLKVAFAIAIVSGSALA.....DSSVMRYQVFGNNATANY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 125546 seqs, 13197846 residues

Total number of hits satisfying chosen parameters: 125546

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents, AA, New:*

1: /cgn2_6/ptodata/1/paa/PCT_NEW_COMB.pep:*

2: /cgn2_6/ptodata/1/paa/US06_NEW_COMB.pep:*

3: /cgn2_6/ptodata/1/paa/US07_NEW_COMB.pep:*

4: /cgn2_6/ptodata/1/paa/US08_NEW_COMB.pep:*

5: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep:*

6: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:*

7: /cgn2_6/ptodata/1/paa/US60_NEW_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	530	68.4	151	US-09-741-873C-4	Sequence 4, Appl
2	452	58.3	131	US-09-741-873C-2	Sequence 2, Appl
3	90.5	11.7	1327	PCT-US04-02338-49	Sequence 49, Appl
4	81	10.5	382	US-10-771-241-299	Sequence 299, Appl
5	79	10.2	878	US-10-771-241-364	Sequence 364, Appl
6	76.5	9.9	321	PCT-US04-05654-550	Sequence 550, Appl
7	76.5	9.9	386	US-10-100-683-10326	Sequence 10326, A
8	76.5	9.9	386	US-10-100-683-10327	Sequence 10327, A
9	76.5	9.9	443	US-10-100-683-7608	Sequence 7608, Ap
10	76	9.8	250	US-10-779-461-40	Sequence 40, Appl
11	75	9.7	205	US-10-767-701-40104	Sequence 40104, A
12	75	9.7	424	US-10-045-674A-551	Sequence 551, Appl
13	75	9.7	533	US-10-045-674A-552	Sequence 527, Appl
14	74	9.5	588	PCT-US04-02188-187	Sequence 187, Appl
15	74	9.5	588	US-10-764-425-187	Sequence 187, Appl
16	73.5	9.5	234	US-10-767-701-45603	Sequence 45603, A
17	73	9.4	246	US-10-779-461-41	Sequence 41, Appl
18	73	9.4	358	US-10-767-701-40444	Sequence 40444, A
19	72	9.3	434	US-10-045-674A-554	Sequence 594, Appl
20	71.5	9.2	379	US-10-767-701-42696	Sequence 42696, A
21	71.5	9.2	386	US-10-767-701-42417	Sequence 42417, A
22	71.5	9.2	1160	US-10-603-150-2	Sequence 2, Appl
23	71	9.2	251	US-10-417-884A-5242	Sequence 5242, Ap
24	71	9.2	364	PCT-US04-05654-1756	Sequence 1756, Ap
25	71	9.2	485	PCT-US04-05654-1844	Sequence 1844, Ap
26	70.5	9.1	251	US-10-767-701-32417	Sequence 32417, A

27	70.5	9.1	273	6	US-10-627-556-392	Sequence 392, Appl
28	70.5	9.1	273	6	US-10-627-556-396	Sequence 396, Appl
29	70.5	9.1	505	6	US-10-627-556-400	Sequence 400, Appl
30	70.5	9.1	505	6	US-10-627-556-402	Sequence 402, Appl
31	70.5	9.1	521	6	US-10-627-556-398	Sequence 398, Appl
32	70.5	9.1	956	6	US-10-093-037A-63	Sequence 63, Appl
33	70.5	9.1	1048	6	PCT-US04-05654-2086	Sequence 2086, Ap
34	70	9.0	175	6	US-10-767-701-35046	Sequence 35046, A
35	70	9.0	199	6	US-10-767-701-51583	Sequence 51583, A
36	70	9.0	245	6	US-10-779-461-59	Sequence 59, Appl
37	70	9.0	399	1	PCT-US04-05654-2748	Sequence 2748, Ap
38	70	9.0	447	6	US-10-767-701-45219	Sequence 45219, A
39	70	9.0	624	1	PCT-US04-05654-1466	Sequence 1466, Ap
40	69.5	9.0	351	1	PCT-US04-05654-1755	Sequence 1755, Ap
41	69	8.9	250	6	US-10-779-461-6	Sequence 6, Appl
42	69	8.9	1871	1	PCT-US03-02038-26	Sequence 26, Appl
43	68.5	8.8	191	6	US-10-767-701-47075	Sequence 47075, A
44	68.5	8.8	251	6	US-10-779-461-30	Sequence 30, Appl
45	68	8.8	175	6	US-10-767-701-35665	Sequence 59665, A

ALIGNMENTS

RESULT 1

US-09-741-873C-4

Sequence 4, Application US/09741873C

GENERAL INFORMATION:

APPLICANT: Normark, Staffan

APPLICANT: Olsen, Arne

TITLE OF INVENTION: Fibrinectin Binding Protein As Well As Its Preparation

FILE REFERENCE: 012889-084

CURRENT APPLICATION NUMBER: US/09/741,873C

CURRENT FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: SE 8801723-1

PRIOR FILING DATE: 1998-05-06

PRIOR APPLICATION NUMBER: US 08/978,878

PRIOR FILING DATE: 1997-11-26

PRIOR APPLICATION NUMBER: US 07/347,189

PRIOR FILING DATE: 1989-05-04

PRIOR APPLICATION NUMBER: US 07/789,437

PRIOR FILING DATE: 1991-11-06

PRIOR APPLICATION NUMBER: US 07/970,846

PRIOR FILING DATE: 1992-11-03

PRIOR APPLICATION NUMBER: US 08/187,865

PRIOR FILING DATE: 1994-01-28

PRIOR APPLICATION NUMBER: US 08/318,519

PRIOR FILING DATE: 1994-10-05

NUMBER OF SEQ ID NOS: 11

SOFTWARE: PatentIn version 3.0

SEQ ID NO 4

LENGTH: 151

TYPE: PRT

ORGANISM: Escherichia coli

US-09-741-873C-4

Query Match Best Local Similarity 68.4%; Score 530; DB 5; Length 151;

Matches 105; Conservative 19; Mismatches 27; Indels 0; Gaps 0;

QY 1 MLLKVAFAIAIVSGSALAGVPPQGGGNGGNGSSGPDSTLSIYGGSNALALQ 60

DB 61 TPARNSDLITTHGGGNGADVGQSDSSIDLTHGGFGNSATLDQWNGNSEVTYKQFGG 120

QY 121 HEMAHANOTASDSSVMRYQVFGNNATANY 151

DB 121 GNGAAVDOTASNSVMRYQVFGNNATANY 151

RESULT 2

```

US-09-741-873C-2
; Sequence 2, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US/09/741,873C
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873C-2

```

```

Query Match
Best Local Similarity 58.3%; Score 452; DB 5; Length 131;
Pred. No. 6,9e-34;
Matches 87; Conservative 18; Mismatches 26; Indels 0; Gaps 0;

```

```

QY 21 GVVPQWGGNNHGGNSGPDSTLSIYQGSANAALALQSDARKSETTITQSGYGNAD 80
DB 1 GVVPQWGGNNHGGNSGPDSTLSIYQGSANAALALQSDARKSETTITQSGYGNAD 60
QY 81 VGGGADNSTIELTONGFRNNATIDQWNAKNDQLVTRVTHMAHANOQASDSSVMRYQV 140
DB 61 VGGGADNSTIELTONGFRNNATIDQWNAKNDQLVTRVTHMAHANOQASDSSVMRYQV 120
QY 141 GFGNNATANQY 151
DB 121 GFGNNATANQY 131

```

RESULT 3

```

PCT-US04-02338-49
; Sequence 49, Application PC/TUS0402338
; GENERAL INFORMATION:
; APPLICANT: EXELIXIS, INC.
; TITLE OF INVENTION: NARPKAS AS MODIFIERS OF THE APC AND AXIN PATHWAYS AND METHODS OF
; FILE REFERENCE: EX04-003C-PC
; CURRENT FILING DATE: 2004-01-28
; PRIOR APPLICATION NUMBER: PCT/US04/02338
; PRIOR FILING DATE: 2004-01-28
; PRIOR APPLICATION NUMBER: US60/443,484
; PRIOR FILING DATE: 2003-01-29
; PRIOR APPLICATION NUMBER: US60/447,358
; PRIOR FILING DATE: 2003-02-11
; PRIOR APPLICATION NUMBER: US60/461,789
; PRIOR FILING DATE: 2003-04-10
; PRIOR APPLICATION NUMBER: US60/470,684
; PRIOR FILING DATE: 2003-05-14
; PRIOR APPLICATION NUMBER: US60/479,650
; PRIOR FILING DATE: 2003-06-19
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 49
; LENGTH: 1327

```

```

; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US04-02338-49

```

```

Query Match
Best Local Similarity 11.7%; Score 90.5; DB 1; Length 1327;
Pred. No. 2.7;
Matches 42; Conservative 16; Mismatches 61; Indels 29; Gaps 6;

```

```

QY 6 VAAFAAI-VVSGALAGVQWGGNNHGGNSGPDSTLSIYQGSANAALALQSDAR 64
DB 99 VAAAPVPAVSTSSAGAVAPNPAGSSNNSPSSSTSS-SSSPSSGSLASPEAA 157
QY 65 KSETTIT---QSGYGNADVGGADNSTIELTONGFRNNATIDQWNAKNDQLVTRVYT 120
DB 158 GVSSTAPLPGAGPGTGVPAVSGALRELLCACRNG-----DVSRYGLVDA--- 204
QY 121 HEMAHANQSD-----SSVMRYQVFG 143
DB 205 -----ANVNAKMDAGKSSPLHPAAGFG 227

```

RESULT 4

```

US-10-771-241-299
; Sequence 299, Application US/10771241
; GENERAL INFORMATION:
; APPLICANT: Zykkind, Judith
; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
; FILE REFERENCE: ELITRA.001C1
; CURRENT FILING DATE: 2004-02-03
; PRIOR APPLICATION NUMBER: US/10/771,241
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: 09/492,709
; PRIOR FILING DATE: 1999-01-27
; NUMBER OF SEQ ID NOS: 485
; SOFTWARE: FaastSeq for Windows Version 3.0
; SEQ ID NO 299
; LENGTH: 382
; TYPE: PRT
; ORGANISM: E. Coli
US-10-771-241-299

```

```

Query Match
Best Local Similarity 10.5%; Score 81; DB 6; Length 382;
Pred. No. 4.1;
Matches 38; Conservative 20; Mismatches 50; Indels 56; Gaps 8;

```

```

QY 30 GNNHGGNSGPDSTLSI-YQ--YGSANAALALQSDARKSETTITQSG-----YNGADYV 82
DB 158 GNTGAGNTSSGTSYSSLANRYGAVNTVGYSRGDS--SQIYYGSGGIAHADITRG 215
QY 83 Q-----GADNSTIE-----LTONGFRNNAT 103
DB 216 QPLGDTNVLVKAPRADVKLENQGTGHTDWRGYAILPFAVEYRENNVALNANSLADNVEL 275
QY 104 DQWNAKNDQLVTRVTH-EMAHANQASDSSVMRYQVFGNNA 146
DB 276 D-----ETVVIVIPTHGALATFNNAQIGGAVMTLTKYKKS 312

```

RESULT 5

```

US-10-771-241-364
; Sequence 364, Application US/10771241
; GENERAL INFORMATION:
; APPLICANT: Zykkind, Judith
; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
; FILE REFERENCE: ELITRA.001C1
; CURRENT FILING DATE: US/10/771,241
; PRIOR APPLICATION NUMBER: 09/492,709

```


; PRIOR FILING DATE: 2000-01-27
 ; PRIOR APPLICATION NUMBER: 60/117,405
 ; PRIOR FILING DATE: 1999-01-27
 ; NUMBER OF SEQ ID NOS: 485
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 364
 ; LENGTH: 878
 ; TYPE: PRT
 ; ORGANISM: E. coli
 ; US-10-771-241-364

Query Match 10.2%; Score 79; DB 6; Length 878;
 Best Local Similarity 24.4%; Pred. No. 17;
 Matches 30; Conservative 16; Mismatches 55; Indels 22; Gaps 3;

QY 49 OXSNAAALALQSDARKSETTI-----TSGYGNAGVGGADNSTIEL 92
 DB 613 QMRHSAASYSMSHDLNGRNTNLAGVYGTILEDNNLSYSTGTGAGGD-GNCGSTGYATL 671
 QY 93 TONGFRNNTIDQWNAKNVDLVTRVVTHEMAHAN-----QTASDSVVRQVFGNNAT 147
 DB 672 NVRGGYGNANIGYSHSDIKQLYGVGGLAHANGVTIGQPLNDTVLVKAPGAKDAKV 731
 QY 148 ANQ 150
 DB 732 ENQ 734

RESULT 6
 PCT-US04-05654-590
 ; Sequence 590, Application PC/TUS0405654
 ; GENERAL INFORMATION:
 ; APPLICANT: Sherman, Bradley K
 ; APPLICANT: Riechmann, Jose Luis
 ; APPLICANT: Jiang, Cai-Zhong
 ; APPLICANT: Heard, Jacqueline E
 ; APPLICANT: Haake, Volker
 ; APPLICANT: Creelman, Robert A
 ; APPLICANT: Ratcliffe, Oliver
 ; APPLICANT: Adam, Luc J
 ; APPLICANT: Reuber, T. Lynne
 ; APPLICANT: Keddie, James
 ; APPLICANT: Dubell III, Arnold N
 ; APPLICANT: Pineda, Omaira
 ; APPLICANT: Repetti, Peter
 ; APPLICANT: Century, Karen
 ; APPLICANT: Guttersen, Neal
 ; APPLICANT: Yu, Guo-Liang
 ; APPLICANT: Broun, Pierre E
 ; APPLICANT: Kumamoto, Roderick W
 ; APPLICANT: Pilgrim, Marsha L
 ; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES IN PLANTS
 ; FILE REFERENCE: MBI-0047 PCT
 ; CURRENT APPLICATION NUMBER: PCT/US04/05654
 ; CURRENT FILING DATE: 2004-03-03
 ; PRIOR APPLICATION NUMBER: 10/374,780
 ; PRIOR FILING DATE: 2003-02-25
 ; PRIOR APPLICATION NUMBER: 10/675,852
 ; PRIOR FILING DATE: 2003-09-30
 ; NUMBER OF SEQ ID NOS: 2950
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 590
 ; LENGTH: 321
 ; TYPE: PRT
 ; ORGANISM: Oryza sativa
 ; FEATURE:
 ; OTHER INFORMATION: G3384 Orthologous to G256
 ; PCT-US04-05654-590

Query Match 9.9%; Score 76.5; DB 1; Length 321;
 Best Local Similarity 22.5%; Pred. No. 8.2;
 Matches 32; Conservative 19; Mismatches 42; Indels 49; Gaps 5;

QY 10 AAIIVSGSLAGVWQW---GGGNHNGGSSGPDSTISIYQSNAAALALQSDARKS 66
 DB 199 AAYASSADNIAIALIGWMPGCG---GGGNKGPEAS-----GSTTATTOOQPOCS 248
 QY 67 ETTTQSGYGNADVGGADNSTIELTONGFRNNTIDQWNAKNVDLVTRVVTHEMAHA 126
 DB 249 -----GEGASASASASOSGAAAAAT-----AQT 272
 QY 127 NOTASDSVVRQVFGNNATA 148
 DB 273 PECSSTETSMATGGAGGAPPA 294

RESULT 7
 US-10-100-683-10326
 ; Sequence 10326, Application US/1010683
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen, et al.
 ; TITLE OF INVENTION: Human Secreted Proteins
 ; FILE REFERENCE: PS900
 ; CURRENT APPLICATION NUMBER: US/10/100,683
 ; CURRENT FILING DATE: 2002-03-19
 ; PRIOR APPLICATION NUMBER: US 60/040,162
 ; PRIOR FILING DATE: 1997-03-07
 ; PRIOR APPLICATION NUMBER: US 60/043,576
 ; PRIOR FILING DATE: 1997-04-11
 ; PRIOR APPLICATION NUMBER: US 60/047,601
 ; PRIOR FILING DATE: 1997-05-23
 ; PRIOR APPLICATION NUMBER: US 60/056,845
 ; PRIOR FILING DATE: 1997-08-22
 ; PRIOR APPLICATION NUMBER: US 60/043,580
 ; PRIOR FILING DATE: 1997-04-11
 ; PRIOR APPLICATION NUMBER: US 60/047,599
 ; PRIOR FILING DATE: 1997-05-23
 ; PRIOR APPLICATION NUMBER: US 60/056,664
 ; PRIOR FILING DATE: 1997-08-22
 ; PRIOR APPLICATION NUMBER: US 60/043,314
 ; PRIOR FILING DATE: 1997-04-11
 ; PRIOR APPLICATION NUMBER: US 60/047,632
 ; PRIOR FILING DATE: 1997-05-23
 ; PRIOR APPLICATION NUMBER: US 60/056,892
 ; PRIOR FILING DATE: 1997-08-22
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 13468
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 10326
 ; LENGTH: 386
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-10-100-683-10326

Query Match 9.9%; Score 76.5; DB 6; Length 386;
 Best Local Similarity 31.1%; Pred. No. 10;
 Matches 28; Conservative 13; Mismatches 34; Indels 15; Gaps 5;
 QY 21 GVVPQ---WGGGNGNNGGSSGPDSTISIYQSNAAALALQSDARKSETTIQSGY 75
 DB 185 GMPGGAHWGCG--NGGPPNFGTITQGAIVAPGVSARA-----SNQNEGCTNPPSGS 237
 QY 76 GNGAD---VGGADNSTIELTONGFRNNT 102
 DB 238 GGGSSNGGSGSGSGSSGSGSGSGSNDNNNGS 267

RESULT 8
 US-10-100-683-10327
 ; Sequence 10327, Application US/1010683
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen, et al.
 ; TITLE OF INVENTION: Human Secreted Proteins
 ; FILE REFERENCE: PS900
 ; CURRENT APPLICATION NUMBER: US/10/100,683
 ; CURRENT FILING DATE: 2002-03-19

```
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 7608
; LENGTH: 443
; TYPE: PRT
; ORGANISM: Homo sapiens
US-100-683-7608

Query Match
  9.9%; Score 76.5; DB 6; Length 443;
Best Local Similarity 31.1%; Pred. No. 12;
Matches 28; Conservative 13; Mismatches 34; Indels 15; Gaps 5;

; PRIOR APPLICATION NUMBER: US 60/040,162
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: US 60/043,576
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: US 60/047,601
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: US 60/056,845
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/043,580
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: US 60/047,599
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: US 60/056,664
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/043,314
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: US 60/047,632
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: US 60/056,892
; PRIOR FILING DATE: 1997-08-22
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 13468
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 10327
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Homo sapiens
US-100-683-10327
```

```
Query Match
  9.9%; Score 76.5; DB 6; Length 386;
Best Local Similarity 31.1%; Pred. No. 10;
Matches 28; Conservative 13; Mismatches 34; Indels 15; Gaps 5;

; PRIOR APPLICATION NUMBER: US 60/040,162
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: US 60/043,576
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: US 60/047,601
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: US 60/056,845
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/043,580
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: US 60/047,599
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: US 60/056,664
; PRIOR FILING DATE: 1997-08-22
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 13468
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 10327
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Homo sapiens
US-100-683-10327
```

```
RESULT 9
US-100-683-7608
; Sequence 7608, Application US/1010683
; GENERAL INFORMATION:
; APPLICANT: Rosen, et al.
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PS900
; CURRENT APPLICATION NUMBER: US/10/100,683
; CURRENT FILING DATE: 2002-03-19
; PRIOR APPLICATION NUMBER: US 60/040,162
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: US 60/043,576
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: US 60/047,601
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: US 60/056,845
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/043,580
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: US 60/047,599
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: US 60/056,664
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/043,314
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: US 60/047,632
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: US 60/056,892
; PRIOR FILING DATE: 1997-08-22
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 13468
```

```
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 7608
; LENGTH: 443
; TYPE: PRT
; ORGANISM: Homo sapiens
US-100-683-7608

Query Match
  9.9%; Score 76.5; DB 6; Length 443;
Best Local Similarity 31.1%; Pred. No. 12;
Matches 28; Conservative 13; Mismatches 34; Indels 15; Gaps 5;

; PRIOR APPLICATION NUMBER: US 60/040,162
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: US 60/043,576
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: US 60/047,601
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: US 60/056,845
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/043,580
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: US 60/047,599
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: US 60/056,664
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/043,314
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: US 60/047,632
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: US 60/056,892
; PRIOR FILING DATE: 1997-08-22
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 13468
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 10327
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Homo sapiens
US-100-683-10327
```

```
Query Match
  9.9%; Score 76.5; DB 6; Length 386;
Best Local Similarity 31.1%; Pred. No. 10;
Matches 28; Conservative 13; Mismatches 34; Indels 15; Gaps 5;

; PRIOR APPLICATION NUMBER: US 60/040,162
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: US 60/043,576
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: US 60/047,601
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: US 60/056,845
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/043,580
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: US 60/047,599
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: US 60/056,664
; PRIOR FILING DATE: 1997-08-22
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 13468
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 10327
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Homo sapiens
US-100-683-10327
```

```
RESULT 10
US-10-779-461-40
; Sequence 40, Application US/10779461
; GENERAL INFORMATION:
; APPLICANT: Morton, Philip A
; TITLE OF INVENTION: ANTIBODIES TO C-MET FOR THE TREATMENT OF CANCERS
; FILE REFERENCE: 00980/1
; CURRENT APPLICATION NUMBER: US/10/779,461
; CURRENT FILING DATE: 2004-02-13
; PRIOR APPLICATION NUMBER: 60/447,073
; PRIOR FILING DATE: 2003-02-13
; NUMBER OF SEQ ID NOS: 161
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 40
; LENGTH: 250
; TYPE: PRT
; ORGANISM: artificial
; FEATURE:
; OTHER INFORMATION: phase display generated human antibody
US-10-779-461-40
```

```
Query Match
  9.8%; Score 76; DB 6; Length 250;
Best Local Similarity 30.6%; Pred. No. 6.7;
Matches 26; Conservative 8; Mismatches 31; Indels 20; Gaps 3;
```

```
; PRIOR APPLICATION NUMBER: US 60/040,162
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: US 60/043,576
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: US 60/047,601
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: US 60/056,845
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/043,580
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: US 60/047,599
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: US 60/056,664
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/043,314
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: US 60/047,632
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: US 60/056,892
; PRIOR FILING DATE: 1997-08-22
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 13468
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 10327
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Homo sapiens
US-100-683-10327
```

```
RESULT 11
US-10-767-701-40104
; Sequence 40104, Application US/10767701
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 40104
; LENGTH: 205
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28WAY03-C44145_1.pep
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:33:51 ; Search time 171.3 Seconds
(without alignments)
860.386 Million cell updates/sec

Title: US-09-543-407-26

Perfect score: 782
Sequence: 1 MLLKVAAPAAIVSGSALA.....DSSVMVROYFGNNATANOY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 6019581 seqs, 976053577 residues

Total number of hits satisfying chosen parameters: 6019581

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Pending Patents_AA_Main:*

1: /cgn2_6/ptodata/2/paa/PCTUS_COMB.pep:*

2: /cgn2_6/ptodata/2/paa/US06_COMB.pep:*

3: /cgn2_6/ptodata/2/paa/US07_COMB.pep:*

4: /cgn2_6/ptodata/2/paa/US08_COMB.pep:*

5: /cgn2_6/ptodata/2/paa/US081_COMB.pep:*

6: /cgn2_6/ptodata/2/paa/US082_COMB.pep:*

7: /cgn2_6/ptodata/2/paa/US083_COMB.pep:*

8: /cgn2_6/ptodata/2/paa/US084_COMB.pep:*

9: /cgn2_6/ptodata/2/paa/US085_COMB.pep:*

10: /cgn2_6/ptodata/2/paa/US086_COMB.pep:*

11: /cgn2_6/ptodata/2/paa/US087_COMB.pep:*

12: /cgn2_6/ptodata/2/paa/US088_COMB.pep:*

13: /cgn2_6/ptodata/2/paa/US089_COMB.pep:*

14: /cgn2_6/ptodata/2/paa/US090_COMB.pep:*

15: /cgn2_6/ptodata/2/paa/US091_COMB.pep:*

16: /cgn2_6/ptodata/2/paa/US092_COMB.pep:*

17: /cgn2_6/ptodata/2/paa/US093_COMB.pep:*

18: /cgn2_6/ptodata/2/paa/US094_COMB.pep:*

19: /cgn2_6/ptodata/2/paa/US095_COMB.pep:*

20: /cgn2_6/ptodata/2/paa/US096_COMB.pep:*

21: /cgn2_6/ptodata/2/paa/US097A_COMB.pep:*

22: /cgn2_6/ptodata/2/paa/US097B_COMB.pep:*

23: /cgn2_6/ptodata/2/paa/US098_COMB.pep:*

24: /cgn2_6/ptodata/2/paa/US099A_COMB.pep:*

25: /cgn2_6/ptodata/2/paa/US099B_COMB.pep:*

26: /cgn2_6/ptodata/2/paa/US100_COMB.pep:*

27: /cgn2_6/ptodata/2/paa/US101_COMB.pep:*

28: /cgn2_6/ptodata/2/paa/US102_COMB.pep:*

29: /cgn2_6/ptodata/2/paa/US103_COMB.pep:*

30: /cgn2_6/ptodata/2/paa/US104_COMB.pep:*

31: /cgn2_6/ptodata/2/paa/US106_COMB.pep:*

32: /cgn2_6/ptodata/2/paa/US107_COMB.pep:*

33: /cgn2_6/ptodata/2/paa/US60_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
------------	-------	-------------	--------	----	-------------

1	782	100.0	151	19	US-09-543-407-26	Sequence 26, Appl
2	709	90.7	151	19	US-09-543-407-20	Sequence 20, Appl
3	692	88.5	151	19	US-09-543-407-5	Sequence 5, Appl
4	687	87.9	151	6	US-08-233-642A-57	Sequence 57, Appl
5	675	86.3	151	19	US-09-543-407-18	Sequence 18, Appl
6	617	78.9	151	19	US-09-543-407-28	Sequence 28, Appl
7	614	78.5	151	19	US-09-543-407-12	Sequence 12, Appl
8	612	78.3	151	19	US-09-543-407-14	Sequence 14, Appl
9	609	77.9	151	19	US-09-543-407-24	Sequence 24, Appl
10	605	77.4	131	19	US-09-543-407-31	Sequence 31, Appl
11	601	76.9	151	19	US-09-543-407-22	Sequence 22, Appl
12	600	76.7	151	19	US-09-543-407-30	Sequence 30, Appl
13	577	73.8	151	19	US-09-543-407-16	Sequence 16, Appl
14	523	66.9	151	19	US-09-543-407-7	Sequence 7, Appl
15	520	66.5	151	13	US-08-978-878-4	Sequence 4, Appl
16	520	66.5	151	21	US-09-741-873B-4	Sequence 4, Appl
17	518	66.2	151	33	US-60-352-946-2	Sequence 2, Appl
18	518	66.2	151	33	US-60-444-371-2	Sequence 2, Appl
19	507	64.8	120	6	US-08-233-642A-55	Sequence 55, Appl
20	473	60.5	109	19	US-09-543-407-34	Sequence 34, Appl
21	466	59.6	158	16	US-09-252-691C-5834	Sequence 5834, Ap
22	466	59.6	158	16	US-09-252-691C-5834	Sequence 5834, Ap
23	466	59.6	158	30	US-10-417-886-5834	Sequence 5834, Ap
24	442	56.5	131	13	US-08-978-878-2	Sequence 2, Appl
25	442	56.5	131	21	US-08-741-873B-2	Sequence 2, Appl
26	338	43.2	109	19	US-09-543-407-35	Sequence 35, Appl
27	266	34.0	68	19	US-09-543-407-37	Sequence 37, Appl
28	201.5	25.8	70	19	US-09-543-407-32	Sequence 32, Appl
29	156	19.9	48	19	US-09-543-407-39	Sequence 39, Appl
30	116.5	14.9	186	16	US-09-252-691-5833	Sequence 5833, Ap
31	116.5	14.9	186	16	US-09-252-691C-5833	Sequence 5833, Ap
32	116.5	14.9	186	30	US-10-417-886-5833	Sequence 5833, Ap
33	116.5	14.9	445	29	US-10-369-493-20638	Sequence 20638, A
34	116.5	14.9	445	33	US-60-360-039-20638	Sequence 20638, A
35	114.5	14.6	145	21	US-09-739-449-8854	Sequence 8854, Ap
36	114.5	14.6	145	23	US-09-803-110-8854	Sequence 8854, Ap
37	112	14.3	151	19	US-09-543-407-8	Sequence 8, Appl
38	107	13.7	151	19	US-09-543-407-6	Sequence 6, Appl
39	105	13.4	597	1	PCT-US01-05992-146	Sequence 146, App
40	105	13.4	597	22	US-09-793-306-146	Sequence 199, App
41	105	13.4	943	1	PCT-US99-03265-199	Sequence 204, App
42	100	12.8	943	1	PCT-US99-03268-204	Sequence 199, App
43	100	12.8	943	13	US-08-942-341-199	Sequence 204, App
44	100	12.8	943	13	US-08-942-341-199	Sequence 199, App
45	100	12.8	943	14	US-09-024-753-199	Sequence 199, App

ALIGNMENTS

RESULT 1
US-09-543-407-26
Sequence 26, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collins, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543,407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 26
LENGTH: 151
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b atga
OTHER INFORMATION: encoding containing the replacement fragment
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.

US-09-543-407-26

Query Match 100.0%; Score 782; DB 19; Length 151;
Best Local Similarity 100.0%; Pred. No. 1.8e-75;
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLKVAFPAIVVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYQGSANALAYDQ 60
DB 1 MKLKVAFPAIVVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYQGSANALAYDQ 60
QY 61 LVTRVVTHEMAHAGNGADVQGGADNSTIELTONGFRNNATIDOMNAKNSDITVQYGG 120
DB 61 LVTRVVTHEMAHAGNGADVQGGADNSTIELTONGFRNNATIDOMNAKNSDITVQYGG 120
QY 121 NNAALVNOTASDSSVWVRQVFGNNATANY 151
DB 121 NNAALVNOTASDSSVWVRQVFGNNATANY 151

RESULT 2

US-09-543-407-20

/ Sequence 20, Application US/09543407
/ GENERAL INFORMATION:
/ APPLICANT: White, Aaron P.
/ APPLICANT: Doran, James L.
/ APPLICANT: Collinson, S. Karen
/ APPLICANT: Kay, William W.
/ TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
/ FILE REFERENCE: 920043.406
/ CURRENT APPLICATION NUMBER: US/09/543.407
/ CURRENT FILING DATE: 2000-04-05
/ NUMBER OF SEQ ID NOS: 59
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 20
/ LENGTH: 151
/ TYPE: PRF
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
/ OTHER INFORMATION: sequence containing the replacement fragment
/ OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-20

Query Match 90.7%; Score 709; DB 19; Length 151;
Best Local Similarity 89.9%; Pred. No. 1.3e-67;
Matches 143; Conservative 0; Mismatches 0; Indels 16; Gaps 2;

QY 1 MKLKVAFPAIVVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYQGSANALAYDQ 57
DB 1 MKLKVAFPAIVVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYQGSANALAYDQ 60
QY 58 -----YDOLVTRVVTHEMAHAGNGADVQGGADNSTIELTONGFRNNATIDOMNAKNSD 112
DB 61 SDARKSETTITQGYGNGADVQGGADNSTIELTONGFRNNATIDOMNAKNSD 112
QY 113 ITVGQYGGNNAALVNOTASDSSVWVRQVFGNNATANY 151
DB 113 ITVGQYGGNNAALVNOTASDSSVWVRQVFGNNATANY 151

RESULT 3

US-09-543-407-5

/ Sequence 5, Application US/09543407
/ GENERAL INFORMATION:
/ APPLICANT: White, Aaron P.
/ APPLICANT: Doran, James L.
/ APPLICANT: Collinson, S. Karen
/ APPLICANT: Kay, William W.
/ TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
/ FILE REFERENCE: 920043.406
/ CURRENT APPLICATION NUMBER: US/09/543.407

CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5
LENGTH: 151
TYPE: PRF
ORGANISM: Salmonella enteritidis
US-09-543-407-5

Query Match 88.5%; Score 692; DB 19; Length 151;
Best Local Similarity 90.7%; Pred. No. 8.7e-66;
Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

QY 1 MKLKVAFPAIVVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYQGSANALAYDQ 60
DB 1 MKLKVAFPAIVVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYQGSANALAYDQ 60
QY 61 LVTRVVTHEMAHAGNGADVQGGADNSTIELTONGFRNNATIDOMNAKNSDITVQYGG 120
DB 61 SDARKSETTITQGYGNGADVQGGADNSTIELTONGFRNNATIDOMNAKNSDITVQYGG 120
QY 121 NNAALVNOTASDSSVWVRQVFGNNATANY 151
DB 121 NNAALVNOTASDSSVWVRQVFGNNATANY 151

RESULT 4

US-08-233-642A-57

/ Sequence 57, Application US/08233642A
/ GENERAL INFORMATION:
/ APPLICANT: Kay, William W.
/ APPLICANT: Collinson, S. Karen
/ APPLICANT: Clouthier, Sharon C.
/ APPLICANT: Doran, James L.
/ TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
/ NUMBER OF SEQUENCES: 58
/ NUMBER OF SEQUENCES: -
/ CORRESPONDENCE ADDRESS:
/ ADDRESS: Seed and Berry
/ STREET: 6300 Columbia Center, 701 Fifth Avenue
/ CITY: Seattle
/ STATE: Washington
/ COUNTRY: U.S.A.
/ ZIP: 98104-7092
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ FILING DATE: 26-APR-1994
/ CLASSIFICATION: 424
/ ATTORNEY/AGENT INFORMATION:
/ NAME: King, Joshua
/ REGISTRATION NUMBER: 35,570
/ REFERENCE/DOCKET NUMBER: 920043.403C3
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (206) 622-4900
/ TELEFAX: (206) 682-6031
/ TELEX: 3723836 SEDANBERY
/ INFORMATION FOR SEQ ID NO: 57:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 151 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
US-08-233-642A-57

Query Match 87.9%; Score 687; DB 6; Length 151;
Best Local Similarity 90.1%; Pred. No. 3e-65;
Matches 136; Conservative 2; Mismatches 13; Indels 0; Gaps 0;

```

QY      1 MLLKVAAPFAIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
      1 MLLKVAAPFAIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
DB      61 LTRVVVTHEMAHAGYNGADVGGADNSTIELTONGFRNNATTIDQWAKNSDITVQYGG 120
      61 SARKESTTTTOSGYNGADVGGADNSTIELTONGFRNNATTIDQWAKNSDITVQYGG 120
QY      121 NNAALVNOTASDSSVMVROVGFNNATANOY 151
      121 NNAALVNOTASDSSVMVROVGFNNATANOY 151
DB      121 NNAALVNOTASDSSVMVROVGFNNATANOY 151

```

RESULT 5

```

US-09-543-407-18
; Sequence 18, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 92043.406
; CURRENT APPLICATION NUMBER: US/09/543.407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-18

```

```

Query Match      86.3%; Score 675; DB 19; Length 151;
Best Local Similarity 81.9%; Pred. No. 5.9e-64;
Matches 136; Conservative 0; Mismatches 0; Indels 30; Gaps 2;

QY      1 MLLKVAAPFAIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
      1 MLLKVAAPFAIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
DB      61 LTRVVVTHEMAHAGYNGADVGGADNSTIELTONGFRNNATTIDQ 105
      61 LTRVVVTHEMAHAGYNGADVGGADNSTIELTONGFRNNATTIDQ 105
QY      46 LTRVVVTHEMAHAGYNGADVGGADNSTIELTONGFRNNATTIDQ 105
      46 LTRVVVTHEMAHAGYNGADVGGADNSTIELTONGFRNNATTIDQ 105
DB      106 WNAKNSDITVQYGSANALVNOTASDSSVMVROVGFNNATANOY 151
      106 WNAKNSDITVQYGSANALVNOTASDSSVMVROVGFNNATANOY 151
QY      106 WNAKNSDITVQYGSANALVNOTASDSSVMVROVGFNNATANOY 151
      106 WNAKNSDITVQYGSANALVNOTASDSSVMVROVGFNNATANOY 151
DB      106 WNAKNSDITVQYGSANALVNOTASDSSVMVROVGFNNATANOY 151

```

RESULT 6

```

US-09-543-407-28
; Sequence 28, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 92043.406
; CURRENT APPLICATION NUMBER: US/09/543.407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 151

```

```

; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-28

```

```

Query Match      78.9%; Score 617; DB 19; Length 151;
Best Local Similarity 73.6%; Pred. No. 1e-57;
Matches 128; Conservative 0; Mismatches 0; Indels 46; Gaps 2;

```

```

QY      1 MLLKVAAPFAIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
      1 MLLKVAAPFAIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
DB      61 LTRVVVTHEMAHAGYNGADVGGADNSTIELTONGFRNNATTIDQWAKNSDITVQYGG 120
      61 SARKESTTTTOSGYNGADVGGADNSTIELTONGFRNNATTIDQWAKNSDITVQYGG 120
QY      121 NNAALVNOTASDSSVMVROVGFNNATANOY 151
      121 NNAALVNOTASDSSVMVROVGFNNATANOY 151
DB      121 NNAALVNOTASDSSVMVROVGFNNATANOY 151

```

RESULT 7

```

US-09-543-407-12
; Sequence 12, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 92043.406
; CURRENT APPLICATION NUMBER: US/09/543.407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-12

```

```

Query Match      78.5%; Score 614; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 2.2e-57;
Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;

```

```

QY      1 MLLKVAAPFAIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
      1 MLLKVAAPFAIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
DB      61 LTRVVVTHEMAHAGYNGADVGGADNSTIELTONGFRNNATTIDQWAKNSDITVQYGG 120
      61 SARKESTTTTOSGYNGADVGGADNSTIELTONGFRNNATTIDQWAKNSDITVQYGG 120
QY      121 NNAALVNOTASDSSVMVROVGFNNATANOY 151
      121 NNAALVNOTASDSSVMVROVGFNNATANOY 151
DB      121 NNAALVNOTASDSSVMVROVGFNNATANOY 151

```

RESULT 8

```

US-09-543-407-14
; Sequence 14, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.

```

```

; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-14

```

```

Query Match          78.3%; Score 612; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 3.6e-57;
Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;

```

```

QY 1 MKLKVAFPAIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALYDQ 60
DB 1 MKLKVAFPAIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
QY 61 LVTRVVTHEMAHAGYGADVGAGADNSTIELTONGFRNNATIDQWNAKNSDITVQYGG 120
DB 61 SDARKSETTITQSGYGADVGAGADNSTIELTONGFRNNATIDQWNAKNSDITVQYDQ 120
QY 121 NNALVNTASDSSVMVROYVGFNNATANY 151
DB 121 LVTRVVTHEMAHAGYVWVROYVGFNNATANY 151

```

RESULT 9

```

US-09-543-407-24
; Sequence 24, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-24

```

```

Query Match          77.9%; Score 609; DB 19; Length 151;
Best Local Similarity 82.1%; Pred. No. 7.5e-57;
Matches 124; Conservative 4; Mismatches 23; Indels 0; Gaps 0;

```

```

QY 1 MKLKVAFPAIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALYDQ 60
DB 1 MKLKVAFPAIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
QY 61 LVTRVVTHEMAHAGYGADVGAGADNSTIELTONGFRNNATIDQWNAKNSDITVQYGG 120
DB 61 SDARKSETTITQSGYGADVGAGADNSTIELTONGFRNNATIDQWNAKNSDITVQYDQ 120

```

```

QY 121 NNALVNTASDSSVMVROYVGFNNATANY 151
DB 121 HEMAHNTASDSSVMVROYVGFNNATANY 151

```

RESULT 10

```

US-09-543-407-31
; Sequence 31, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Salmonella enteritidis
US-09-543-407-31

```

```

Query Match          77.4%; Score 605; DB 19; Length 131;
Best Local Similarity 89.3%; Pred. No. 1.7e-56;
Matches 117; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

```

```

QY 21 GVPWQGGGNNHNGGNSGPDSTLSIYQGSANALALYDQLVTRVVTHEMAHAGYGAD 80
DB 1 GVPWQGGGNNHNGGNSGPDSTLSIYQGSANALALQSDARKSETTITQSGYGAD 60
QY 81 VQGGADNSTIELTONGFRNNATIDQWNAKNSDITVQYGSANALVNTASDSSVMVROY 140
DB 61 VQGGADNSTIELTONGFRNNATIDQWNAKNSDITVQYGSANALVNTASDSSVMVROY 120
QY 141 GFNNATANY 151
DB 121 GFNNATANY 131

```

RESULT 11

```

US-09-543-407-22
; Sequence 22, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 22
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-22

```

```

Query Match          76.9%; Score 601; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 5.5e-56;
Matches 123; Conservative 4; Mismatches 24; Indels 0; Gaps 0;

```

```

QY 1 MKLKVAFPAIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALYDQ 60

```



```

Db      1 MKLKVAAFAIIVSGSALAGVVPQWGGGNNHGGSSGPDSTLSIYQGSANAALALQ 60
        |||
Qy      61 LVTRVVTHEMAHAGYNGADVGGAGDNSTIELTQNGFRNNATIDQWNAKSDITVGGYGG 120
        |||
Db      61 SDAKSETTITGGYGGAGADVGGAGDNSTIELTQNGFRNNATIDQWNAKSDITVGGYGG 120
        |||
Qy      121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151
        |||
Db      121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151
        |||

```

RESULT 12

```

US-09-543-407-30
; Sequence 30, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543.407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
US-09-543-407-30

```

```

Query Match      76.7%; Score 600; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 7e-56;
Matches 123; Conservative 4; Mismatches 24; Indels 0; Gaps 0;

```

```

Qy      1 MKLKVAAFAIIVSGSALAGVVPQWGGGNNHGGSSGPDSTLSIYQGSANAALALQ 60
        |||
Db      1 MKLKVAAFAIIVSGSALAGVVPQWGGGNNHGGSSGPDSTLSIYQGSANAALALQ 60
        |||
Qy      61 LVTRVVTHEMAHAGYNGADVGGAGDNSTIELTQNGFRNNATIDQWNAKSDITVGGYGG 120
        |||
Db      61 SDAKSETTITGGYGGAGADVGGAGDNSTIELTQNGFRNNATIDQWNAKSDITVGGYGG 120
        |||
Qy      121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151
        |||
Db      121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151
        |||

```

RESULT 13

```

US-09-543-407-16
; Sequence 16, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543.407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence

```

```

; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding P73 from GPe3 of Leishmania major.
US-09-543-407-16

```

```

Query Match      73.8%; Score 577; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 2.1e-53;
Matches 122; Conservative 5; Mismatches 24; Indels 0; Gaps 0;

```

```

Qy      1 MKLKVAAFAIIVSGSALAGVVPQWGGGNNHGGSSGPDSTLSIYQGSANAALALQ 60
        |||
Db      1 MKLKVAAFAIIVSGSALAGVVPQWGGGNNHGGSSGPDSTLSIYQGSANAALALQ 60
        |||
Qy      61 LVTRVVTHEMAHAGYNGADVGGAGDNSTIELTQNGFRNNATIDQWNAKSDITVGGYGG 120
        |||
Db      61 SDAKSETTITGGYGGAGADVGGAGDNSTIELTQNGFRNNATIDQWNAKSDITVGGYGG 120
        |||
Qy      121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151
        |||
Db      121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151
        |||

```

RESULT 14

```

US-09-543-407-7
; Sequence 7, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543.407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-543-407-7

```

```

Query Match      66.9%; Score 523; DB 19; Length 151;
Best Local Similarity 68.9%; Pred. No. 1.4e-47;
Matches 104; Conservative 18; Mismatches 29; Indels 0; Gaps 0;

```

```

Qy      1 MKLKVAAFAIIVSGSALAGVVPQWGGGNNHGGSSGPDSTLSIYQGSANAALALQ 60
        |||
Db      1 MKLKVAAFAIIVSGSALAGVVPQWGGGNNHGGSSGPDSTLSIYQGSANAALALQ 60
        |||
Qy      61 LVTRVVTHEMAHAGYNGADVGGAGDNSTIELTQNGFRNNATIDQWNAKSDITVGGYGG 120
        |||
Db      61 TDARNSDLITTHGGGNGADVGGAGDNSTIELTQNGFRNNATIDQWNAKSDITVGGYGG 120
        |||
Qy      121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151
        |||
Db      121 GNGAAVDQTAHSNSVMVRQVGFNNATANQY 151
        |||

```

RESULT 15

```

US-08-978-878-4
; Sequence 4, Application US/08978878
; GENERAL INFORMATION:
; APPLICANT: NORMARK, Staffan
; APPLICANT: OLSEN, Arne
; TITLE OF INVENTION: FIBRONECTIN BINDING PROTEIN AS WELL AS ITS PREPARATION
; FILE REFERENCE: 012889-081
; CURRENT APPLICATION NUMBER: US/08/978.878
; CURRENT FILING DATE: 1997-11-26
; EARLIER APPLICATION NUMBER: SE 8801723-1
; EARLIER FILING DATE: 1988-05-06

```

```

; EARLIER APPLICATION NUMBER: US 07/347,189
; EARLIER FILING DATE: 1989-05-04
; EARLIER APPLICATION NUMBER: US 07/789,437
; EARLIER FILING DATE: 1991-11-06
; EARLIER APPLICATION NUMBER: US 07/970,846
; EARLIER FILING DATE: 1992-11-03
; EARLIER APPLICATION NUMBER: US 08/187,865
; EARLIER FILING DATE: 1994-01-28
; EARLIER APPLICATION NUMBER: US 08/318,519
; EARLIER FILING DATE: 1994-10-05
; EARLIER APPLICATION NUMBER: US 08/495,959
; EARLIER FILING DATE: 1995-06-28
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-08-978-878-4

```

```

Query Match      66.5%; Score 520; DB 13; Length 151;
Best Local Similarity 68.2%; Pred. No: 2.9e-47;
Matches 103; Conservative 19; Mismatches 29; Indels 0; Gaps 0;

```

```

QY      1  MKLKVAFPAIIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQYGSANALYDQ 60
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      1  MKLKVAAIAIIVSGSAGVVPQYGGGNNHGGGNSGPNSELNIYQYGGNSALALQ 60
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY      61  LVTRVVTHEMAHAGYGAGADVGGAGDNSTELTONGFRNNATTIDQWNAKNSDITVGOYGG 120
      :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
Db      61  TDANSDLTTOHGGAGADVGGGSDSDSIDLTQRFGNSATLDDQWNGKNSMTVKQFGG 120
      :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||

QY      121  NNALVNOQTASDSSVMRQVGFNNATTNOY 151
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      121  NGGAIVDQTASNSSSVNTQVGFNNATTAHQY 151
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

Search completed: March 11, 2004, 19:13:13
 Job time : 171.3 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:35:05 ; Search time 5.3 Seconds
(without alignments)
376.014 Million cell updates/sec

Title: US-09-543-407-26

Perfect score: 782

Sequence: 1 MKLLKVAAPAAIVSGSALA.....DSSVMRVQVFGNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 125546 seqs, 13197846 residues

Total number of hits satisfying chosen parameters: 125546

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents_AA_New:*
1: /cgn2_6/ptodata/1/paa/PCT_NEW_COMB.pap:*
2: /cgn2_6/ptodata/1/paa/US06_NEW_COMB.pap:*
3: /cgn2_6/ptodata/1/paa/US07_NEW_COMB.pap:*
4: /cgn2_6/ptodata/1/paa/US08_NEW_COMB.pap:*
5: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pap:*
6: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pap:*
7: /cgn2_6/ptodata/1/paa/US60_NEW_COMB.pap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	520	66.5	151	US-09-741-873C-4	Sequence 4, App1
2	442	56.5	131	US-09-741-873C-2	Sequence 2, App1
3	87.5	11.2	234	US-10-767-701-45603	Sequence 45603, A
4	83.5	10.7	443	US-10-100-683-7608	Sequence 7608, Ap
5	82	10.5	1327	PCT-US04-02338-49	Sequence 49, App1
6	81.5	10.4	179	US-10-767-701-35342	Sequence 35342, A
7	78	10.0	351	PCT-US04-05654-336	Sequence 336, App
8	77.5	9.9	366	US-10-100-683-10326	Sequence 10326, A
9	77.5	9.9	366	US-10-100-683-10327	Sequence 10327, A
10	76.5	9.8	348	US-10-004-1158-34	Sequence 34, App1
11	76	9.7	430	US-10-451-467A-314	Sequence 314, App
12	75	9.6	356	US-10-093-037A-63	Sequence 63, App1
13	75	9.6	1160	US-10-603-150-2	Sequence 2, App1
14	73.5	9.4	878	US-10-771-241-364	Sequence 364, App
15	72.5	9.3	273	US-10-627-556-392	Sequence 392, App
16	72.5	9.3	273	US-10-627-556-396	Sequence 396, App
17	72.5	9.3	505	US-10-627-556-400	Sequence 400, App
18	72.5	9.3	505	US-10-627-556-402	Sequence 402, App
19	72.5	9.3	521	US-10-627-556-398	Sequence 398, App
20	72	9.2	199	US-10-767-701-45448	Sequence 45448, A
21	72	9.2	331	PCT-US04-05654-590	Sequence 590, App
22	72	9.2	335	PCT-US04-04499-132	Sequence 132, App
23	72	9.2	588	PCT-US04-02188-187	Sequence 187, App
24	72	9.2	588	US-10-764-425-187	Sequence 187, App
25	71	9.1	863	US-60-544-461-29	Sequence 29, App1
26	70	9.0	191	US-10-767-701-47075	Sequence 47075, A

27	70	9.0	305	1	PCT-US04-05654-1406	Sequence 1406, Ap
28	70	9.0	339	1	PCT-US04-05654-2748	Sequence 2748, Ap
29	70	9.0	434	6	US-10-045-674A-594	Sequence 594, App
30	69	8.8	215	6	US-10-767-701-43375	Sequence 43375, A
31	69	8.8	245	6	US-10-779-461-59	Sequence 59, App1
32	69	8.8	616	6	US-10-467-243-1	Sequence 1, App1
33	68.5	8.8	311	6	US-10-786-850-1	Sequence 1, App1
34	68.5	8.8	424	6	US-10-045-674A-591	Sequence 591, App
35	68.5	8.8	433	1	PCT-US04-05654-2586	Sequence 2586, Ap
36	68.5	8.8	442	1	PCT-US04-05654-338	Sequence 338, App
37	68.5	8.8	533	6	US-10-045-674A-527	Sequence 527, App
38	68	8.7	195	6	US-10-767-701-32144	Sequence 32144, A
39	68	8.7	351	1	PCT-US04-05654-1755	Sequence 1755, Ap
40	68	8.7	363	6	US-10-767-701-37968	Sequence 37968, A
41	68	8.7	636	6	US-10-048-770C-1	Sequence 1, App1
42	67.5	8.6	201	1	PCT-US04-05654-136	Sequence 136, App
43	67.5	8.6	245	6	US-10-767-701-41094	Sequence 41094, A
44	67.5	8.6	358	6	US-10-767-701-40444	Sequence 40444, A
45	67.5	8.6	485	1	PCT-US04-05654-1844	Sequence 1844, Ap

ALIGNMENTS

```

RESULT 1
US-09-741-873C-4
; Sequence 4, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibrinectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741, 873C
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978, 878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347, 189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789, 437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970, 846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187, 865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318, 519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
; US-09-741-873C-4

```

Query Match 66.5%; Score 520; DB 5; Length 151;
Best Local Similarity 68.2%; Pred. No. 1.4e-40;
Matches 103; Conservative 19; Mismatches 29; Indels 0; Gaps 0;

```

QY 1 MKLLKVAAPAAIVSGSALAGVVPQGGGNNHGGNSGPDSTLSIYQGSANALAYDQ 60
    |||||
DB 1 MKLLKVAAPAAIVSGSALAGVVPQGGGNNHGGNSGPDSTLSIYQGSANALAYDQ 60
    |||||
QY 61 LVTRVYTHMAHGYGNGADVGGAGNISTELTQNFRRNATTDDNNARSDITVQOYCG 120
    |||||
DB 61 TDARNSDLITTOHGGNGADVGGSDSSSIDLTOHGFNSATTDQNNGNSEMTVAFQGG 120
    |||||
QY 121 NNAALVNTASDSVVRVQVFGNNATANQY 151
    |||||
DB 121 GNGAAVDOTASNSVVRVQVFGNNATANQY 151
    |||||

```

RESULT 2
US-09-741-873C-2
Sequence 2, Application US/09741873C
GENERAL INFORMATION:
APPLICANT: Normark, Steffan
TITLE OF INVENTION: Fibroectin Binding Protein As Well As Its Preparation
FILE REFERENCE: 012889-084
CURRENT APPLICATION NUMBER: US/09/741,873C
CURRENT FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: SE 8801723-1
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: US 08/978,878
PRIOR FILING DATE: 1997-11-26
PRIOR APPLICATION NUMBER: US 07/347,189
PRIOR FILING DATE: 1989-05-04
PRIOR APPLICATION NUMBER: US 07/789,437
PRIOR FILING DATE: 1991-11-06
PRIOR APPLICATION NUMBER: US 07/970,846
PRIOR FILING DATE: 1992-11-03
PRIOR APPLICATION NUMBER: US 08/187,865
PRIOR FILING DATE: 1994-01-28
PRIOR APPLICATION NUMBER: US 08/318,519
PRIOR FILING DATE: 1994-10-05
NUMBER OF SEQ ID NOS: 11
SOFTWARE: PatentIn version 3.0
SEQ ID NO 2
LENGTH: 131
TYPE: PRT
ORGANISM: Escherichia coli
US-09-741-873C-2

Query Match 56.5%; Score 442; DB 5; Length 131;
Best Local Similarity 64.9%; Pred. No. 1.5e-33;

Matches 85; Conservative 18; Mismatches 28; Indels 0; Gaps 0;

QY 21 GVVPWGGGNNHGGSSGPDSTLSIYQGSANALYDQVTVRVVTHEMAGYNGAD 80
DB 1 GVPVPGGGGNNHGGSSGNNSEINLTQYGGNSALLQTDARNSDLTIHGGNGAD 60
QY 81 VGGADNSTIELTONGFRNNATIDQMANSDITVGYGNNALVNOTASDSSVWVQV 140
DB 61 VGGSDSDSIDLTQRFNSATLTDQNGKNSMTVKQPGGGAADVDTASNSVNVTV 120
QY 141 GFGNNATANOY 151
DB 121 GFGNNATANOY 131

RESULT 3
US-10-767-701-45603
Sequence 45603, Application US/10767701
GENERAL INFORMATION:
APPLICANT: Kovacic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
FILE REFERENCE: 38-21(5353)B
CURRENT APPLICATION NUMBER: US/10/767,701
CURRENT FILING DATE: 2004-01-29
NUMBER OF SEQ ID NOS: 63128
SEQ ID NO 45603
LENGTH: 234
TYPE: PRT
ORGANISM: Sorghum bicolor
FEATURE:
NAME/KEY: unsure
LOCATION: (1) (234)
OTHER INFORMATION: unsure at all Xaa locations
FEATURE:
OTHER INFORMATION: Clone ID: SORBI-28MAY03-C82834_1.pep
US-10-767-701-45603

Query Match 11.2%; Score 87.5; DB 6; Length 234;
Best Local Similarity 23.6%; Pred. No. 0.45;
Matches 38; Conservative 13; Mismatches 45; Indels 65; Gaps 6;

QY 1 MKLKVAFPAIVV--SGSALAGVPPW---GGGNNHGGN-----SSGPDSTL 45
DB 1 MATTKLAALCFIVLIGGANNARVAVYAGGGGGGGGRGRMGASRMSSGSGSC 60
QY 46 SIY-----QYSANALYDQVTVRVVTHEMAGYNGADV 82
DB 61 GXYEAGSGXAVYAGGGGGGGGGGGGGGGG-----SGYSGSGYG 102
QY 83 QGADNSTIELTONGFRNNATIDQMANSDITVGYGNNALV 123
DB 103 QAGSGS-----NGAVYAGGAGGGGGGGGGGGGG 134

RESULT 4
US-10-100-683-7608

Sequence 7608, Application US/10100683
GENERAL INFORMATION:
APPLICANT: Rosen, et al.

TITLE OF INVENTION: Human Secreted Proteins
FILE REFERENCE: P8900

CURRENT APPLICATION NUMBER: US/10/100,683
PRIOR FILING DATE: 2002-03-19
PRIOR APPLICATION NUMBER: US 60/040,162
PRIOR FILING DATE: 1997-03-07
PRIOR APPLICATION NUMBER: US 60/043,576
PRIOR FILING DATE: 1997-04-11
PRIOR APPLICATION NUMBER: US 60/047,601
PRIOR FILING DATE: 1997-05-23
PRIOR APPLICATION NUMBER: US 60/056,845
PRIOR FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: US 60/043,580
PRIOR FILING DATE: 1997-04-11
PRIOR APPLICATION NUMBER: US 60/047,599
PRIOR FILING DATE: 1997-05-23
PRIOR APPLICATION NUMBER: US 60/056,664
PRIOR FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: US 60/043,314
PRIOR FILING DATE: 1997-04-11
PRIOR APPLICATION NUMBER: US 60/047,632
PRIOR FILING DATE: 1997-05-23
PRIOR APPLICATION NUMBER: US 60/056,892
PRIOR FILING DATE: 1997-08-22
REMAINING PRIOR APPLICATION DATA REMOVED - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 13468
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 7608
LENGTH: 443
TYPE: PRT
ORGANISM: Homo sapiens
US-10-100-683-7608

Query Match 10.7%; Score 83.5; DB 6; Length 443;
Best Local Similarity 24.2%; Pred. No. 2.3;
Matches 31; Conservative 20; Mismatches 42; Indels 35; Gaps 6;

QY 21 GVVPQ---WGGGNNHGGSSGPDSTLSIYQ--YGSANALYDQVTVRVVTH---EMA 71
DB 185 GMPFQAPMGQGG--NGPPNFTNTOGAVAQPGYGSVVASNNECTNPPSGSGGSS 242
QY 72 HAGYNGADVQGADNSTIELTONGFRNNATI-----DQMANKSDI 113
DB 243 NSGGSGSGSGSGSGSGS-----NGDNNNGSSSGSGSGSGSGSGSGSGSGS 236
QY 114 TVGYGNN 121
DB 297 SSGNHGGS 304

RESULT 5
PCT-US04-02338-49
; Sequence 49, Application PC/TUS0402338
; GENERAL INFORMATION:
; APPLICANT: EXELIXIS, INC.
; TITLE OF INVENTION: MAPKX AS MODIFIERS OF THE APC AND AXIN PATHWAYS AND METHODS OF
; FILE REFERENCE: EX04-003C-PC
; CURRENT APPLICATION NUMBER: PCT/US04/02338
; PRIOR FILING DATE: 2004-01-28
; PRIOR APPLICATION NUMBER: US60/443,484
; PRIOR FILING DATE: 2003-01-29
; PRIOR APPLICATION NUMBER: US60/447,358
; PRIOR FILING DATE: 2003-02-11
; PRIOR APPLICATION NUMBER: US60/461,789
; PRIOR FILING DATE: 2003-04-10
; PRIOR APPLICATION NUMBER: US60/470,684
; PRIOR FILING DATE: 2003-05-14
; PRIOR APPLICATION NUMBER: US60/479,650
; PRIOR FILING DATE: 2003-06-19
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 49
; LENGTH: 1327
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US04-02338-49

Query Match 10.5%; Score 82; DB 1; Length 1327;
Best Local Similarity 29.6%; Pred. No. 12;
Matches 34; Conservative 14; Mismatches 59; Indels 8; Gaps 4;
QY 6 VAAFAAI-VVSGSALGVVPOWGGGNNHNGGNSGPDSTLSIYQYSANALYDQ--- 61
DB 99 VAAAPVPAVSTISAGAVNANPAGSSGNNPSSSSSTSS-SSSSPSSPSSLAESPEA 157
QY 62 -VTRVVTHEMAHAGYNGADVGGADNSTIELTONG--FNNNATTIDOMAKNSDI.113
DB 158 GVSSTAPLPGGAGPGTGVPAVSGALRELEACRNGDVSHVKLVDAANVNAKDM 212

RESULT 6
US-10-767-701-35342
; Sequence 35342, Application US/10767701
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; PRIOR FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 35342
; LENGTH: 179
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-C65560_1.pcp
US-10-767-701-35342

Query Match 10.4%; Score 81.5; DB 6; Length 179;
Best Local Similarity 25.2%; Pred. No. 1.1;
Matches 32; Conservative 13; Mismatches 47; Indels 35; Gaps 5;
QY 27 GGGGNHNGGNSGPDSTLSIYQYSANALYDQVTRVVTHEMAHAGYNGADVGGAD 86
DB 4 GGGGNHNGGNSGPDSTLSIYQYSANALYDQVTRVVTHEMAHAGYNGADVGGAD 46
QY 87 NSTIELTONGFRNNATTIDOMAKNSDIITVQYGGNNALYDQ-----SSSVVVRQ 139
DB 47 DNT-----SSSSGDDGWSKSS--GGYGGWDEILDNNITAAQAENRSSNSFSAGN 95

QY 140 VGEFNNNA 146
DB 96 NGYGTGS 102

RESULT 7
PCT-US04-05654-336
; Sequence 336, Application PC/TUS0405654
; GENERAL INFORMATION:
; APPLICANT: Sherman, Bradley K.
; APPLICANT: Riechmann, Jose Luis
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Heard, Jacqueline E
; APPLICANT: Haake, Volker
; APPLICANT: Creelman, Robert A
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Adam, Luc J
; APPLICANT: Reuber, T. Lynne
; APPLICANT: Keddle, James
; APPLICANT: Dubeil III, Arnold N
; APPLICANT: Pineda, Omaira
; APPLICANT: Repetti, Peter
; APPLICANT: Century, Karen
; APPLICANT: Gulterson, Neal
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Brown, Pierre E
; APPLICANT: Kumimoto, Roderick W
; APPLICANT: Pilgrim, Marsha L
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES IN PLANTS
; FILE REFERENCE: MBI-0047 PCT
; CURRENT APPLICATION NUMBER: PCT/US04/05654
; PRIOR FILING DATE: 2004-03-03
; PRIOR APPLICATION NUMBER: 10/374,780
; PRIOR FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: 10/675,852
; PRIOR FILING DATE: 2003-09-30
; NUMBER OF SEQ ID NOS: 2950
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 336
; LENGTH: 391
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: G1794 (conserved domain in AA coordinates: 182-249)
PCT-US04-05654-336

Query Match 10.0%; Score 78; DB 1; Length 391;
Best Local Similarity 25.4%; Pred. No. 6.2;
Matches 30; Conservative 23; Mismatches 55; Indels 10; Gaps 4;
QY 38 SSGPDSTLSIYQYSANALYDQVTRVVTHEMAHAGYNGADVGGADNSTIELTONGF 97
DB 57 SSSSSSLGFSREMMSAIVSAL-THVAVGNVPOHQGGGSGSGSTNS--SSSGQ 112
QY 98 RNNATTIDOMAK--NSDITVGOY--GGNNALYDQVTSDDSSVVRVGVFGNNATAN 149
DB 113 KRRREVEEGGAKAVKANTLTVDQYFGSSSTSKVREASNNSGPPTTEYTTATAS 170

RESULT 8
US-10-100-683-10326
; Sequence 10326, Application US/10100683
; GENERAL INFORMATION:
; APPLICANT: Rosen, et al.
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PS900
; CURRENT APPLICATION NUMBER: US/10/100,683
; PRIOR FILING DATE: 2002-03-19
; PRIOR APPLICATION NUMBER: US 60/040,162
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: US 60/043,576
; PRIOR FILING DATE: 1997-04-11

```
; PRIOR APPLICATION NUMBER: US 60/047,601
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: US 60/056,845
; PRIOR FILING DATE: 1997-06-22
; PRIOR APPLICATION NUMBER: US 60/043,580
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: US 60/047,599
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: US 60/056,664
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/043,314
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: US 60/047,632
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: US 60/056,892
; PRIOR FILING DATE: 1997-08-22
; Remaining Prior Application data removed - See File Wrapper or PALM.
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 10326
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-100-683-10326
```

```
Query Match          9.9%; Score 77.5; DB 6; Length 386;
Best Local Similarity 25.6%; Pred. No. 6.8; Indels 23; Gaps 6;
Matches 33; Conservative 21; Mismatches 52;
```

```
QY 21 GVVPQ--WGGGNNHNGGNSGPDSTLSIQ--YGSANAALYDQVTVTVTH----EMA 71
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 185 GMPDQAPWQGG--NGPPNFGTNTQGAVALQPGYGVASNQNEGCTNPPPSGGGSS 242
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 72 HAGYNGADVQGGADNSTIELTONGFRNNATITDQMANKSDI-----TVQYGNNAAL 125
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 243 NSGGSGSGSGSGSGSGS-----NGDNNNGSSSGSGSGSGSGSGSGSGSGSGS 296
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 126 VNQTASDSS 134
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 297 RQDSGSESS 305
```

```
RESULT 9
US-10-100-683-10327
; Sequence 10327, Application US/10100683
; GENERAL INFORMATION:
; APPLICANT: Rosen, et al.
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PS900
; CURRENT APPLICATION NUMBER: US/10/100,683
; CURRENT FILING DATE: 2002-03-19
; PRIOR APPLICATION NUMBER: US 60/040,162
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: US 60/043,576
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: US 60/047,601
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: US 60/056,845
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/043,580
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: US 60/047,599
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: US 60/056,664
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/043,314
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: US 60/047,632
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: US 60/056,892
; PRIOR FILING DATE: 1997-08-22
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 13468
```

```
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 10327
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-100-683-10327
```

```
Query Match          9.9%; Score 77.5; DB 6; Length 386;
Best Local Similarity 25.6%; Pred. No. 6.8; Indels 23; Gaps 6;
Matches 33; Conservative 21; Mismatches 52;
```

```
QY 21 GVVPQ--WGGGNNHNGGNSGPDSTLSIQ--YGSANAALYDQVTVTVTH----EMA 71
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 185 GMPDQAPWQGG--NGPPNFGTNTQGAVALQPGYGVASNQNEGCTNPPPSGGGSS 242
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 72 HAGYNGADVQGGADNSTIELTONGFRNNATITDQMANKSDI-----TVQYGNNAAL 125
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 243 NSGGSGSGSGSGSGSGS-----NGDNNNGSSSGSGSGSGSGSGSGSGSGSGS 296
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 126 VNQTASDSS 134
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 297 RQDSGSESS 305
```

```
RESULT 10
US-10-004-115B-34
; Sequence 34, Application US/10004115B
; GENERAL INFORMATION:
; APPLICANT: ASAKO, HIROYUKI
; APPLICANT: MATSUMURA, KENJI
; APPLICANT: SHIMIZU, MASATOSHI
; APPLICANT: ITO, NOBUYA
; APPLICANT: MAKITA, RYUHEI
; TITLE OF INVENTION: PROCESS FOR PRODUCING OPTICALLY ACTIVE
; FILE REFERENCE: 4-HALO-3-HYDROXYBUTANOLATE
; CURRENT APPLICATION NUMBER: US/10/004,115B
; CURRENT FILING DATE: 2001-12-06
; PRIOR APPLICATION NUMBER: JP 2000-372704
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: JP 2001-006144
; PRIOR FILING DATE: 2001-01-15
; PRIOR APPLICATION NUMBER: JP 2001-026594
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: JP 2001-175175
; PRIOR FILING DATE: 2001-06-11
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: Patent In Ver. 3.2
; SEQ ID NO 34
; LENGTH: 348
; TYPE: PRT
; ORGANISM: Corynebacterium sp.
US-10-004-115B-34
```

```
Query Match          9.8%; Score 76.5; DB 6; Length 348;
Best Local Similarity 27.3%; Pred. No. 7.3;
Matches 30; Conservative 15; Mismatches 42; Indels 23; Gaps 3;
```

```
QY 51 GSANAALYDQVTVTVTHMAHAGYNGADV-----GGGADNSTIELTONGFRNNATITQ 105
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 236 GSQGAALVLDVFGYQPTIDTAMAVAGVSDVITVIGIDQAHAKVFPSPPEASVTVPY 295
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 106 WNAKNS-----DITVGYGNNAA-----LVNQTASDSSVWV 137
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 296 WGARNELIELIDLAHAGIFDISVETPSLDNGAEAYRRLAAGTUSGRAVVV 345
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 11
US-10-451-467A-314
; Sequence 314, Application US/10451467A
; GENERAL INFORMATION:
; APPLICANT: CONTRERAS, ROLAND HENRI
; APPLICANT: EBERHARDT, INES
```

APPLICANT: LUYTEN, WALTER HERMAN MARIA LOUIS
 APPLICANT: REEKMAN, RIEKA JOSEPHINA
 TITLE OF INVENTION: BAX-RESPONSIVE GENES FOR DRUG TARGET IDENTIFICATION IN
 TITLE OF INVENTION: YEAST AND FUNGI
 FILE REFERENCE: JAB-1667
 CURRENT APPLICATION NUMBER: US/10/451,467A
 PRIOR FILING DATE: 2003-06-19
 PRIOR APPLICATION NUMBER: EP 00870318.3
 PRIOR FILING DATE: 2000-12-22
 PRIOR APPLICATION NUMBER: EP 01870002.1
 PRIOR FILING DATE: 2001-01-04
 PRIOR APPLICATION NUMBER: EP 01870003.9
 PRIOR FILING DATE: 2001-01-09
 NUMBER OF SEQ ID NOS: 732
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO 314
 LENGTH: 430
 TYPE: PRT
 ORGANISM: Saccharomyces cerevisiae
 US-10-451-467A-314

Query Match 9.7%; Score 76; DB 6; Length 430;
 Best Local Similarity 25.6%; Pred. No. 11;
 Matches 32; Conservative 15; Mismatches 50; Indels 28; Gaps 6;

QY 29 GGNHGGGNSGPDSTLSTIYOGSANAALYDQVTRVVTHEMAHAGYNGADVGGADN- 87
 DB 148 GSNNDSDYGSNNDS-----YGSNNDSYG-----SSKNKSSYGSNNDSYGSNNDS 194

QY 88 -----STIELTQNGFRNNATIDQNNAKKSDITVQYQYGGNAAALVNGTASDS--SVMMQV 140
 DB 195 DSYGSSNKKSSYGSNN--DSYGSNNDD-----SYGSNNDSYGSNNDSYGSNNKKS 247

QY 141 GFGNN 145
 DB 248 SYGSN 252

RESULT 12
 US-10-093-037A-63
 Sequence 63, Application US/10093037A
 GENERAL INFORMATION:
 APPLICANT: Jay M. Short
 APPLICANT: Bylina, Edward
 APPLICANT: Swanson, Ronald V.
 APPLICANT: Machur, Eric J.
 APPLICANT: Lam, David E.
 TITLE OF INVENTION: ENZYMES HAVING GLYCOSIDASE ACTIVITY AND METHODS OF USE THEREOF
 FILE REFERENCE: 564462001402
 CURRENT APPLICATION NUMBER: US/10/093,037A
 PRIOR FILING DATE: 2002-03-06
 PRIOR APPLICATION NUMBER: US 09/910,579
 PRIOR FILING DATE: 2001-07-20
 PRIOR APPLICATION NUMBER: US 09/134,078
 PRIOR FILING DATE: 1998-08-13
 PRIOR APPLICATION NUMBER: US 08/949,026
 PRIOR FILING DATE: 1997-10-10
 PRIOR APPLICATION NUMBER: US 60/056,916
 PRIOR FILING DATE: 1996-12-06
 NUMBER OF SEQ ID NOS: 72
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 63
 LENGTH: 956
 TYPE: PRT
 ORGANISM: Bankia Gouldi
 US-10-093-037A-63

Query Match 9.6%; Score 75; DB 6; Length 956;
 Best Local Similarity 24.2%; Pred. No. 35;
 Matches 31; Conservative 17; Mismatches 50; Indels 30; Gaps 6;

QY 30 GNNHGGGNSGPDSTLSTIYOGSANAALYDQVTRVVTHEMAHAGYNGADVGGADN- 89
 DB 148 GSNNDSDYGSNNDS-----YGSNNDSYG-----SSKNKSSYGSNNDSYGSNNDS 194

DB 369 GNNPFOKIQCAFNR-SVLYGSANGSTNDEIIR-----GESATNP 410
 QY 90 IEL---TONGFRNNATIDQNNAKKSDITVQYQY-----GNNAAVNOTASD---SSVVR 138
 DB 411 VFGSLDINNGLTSLSEBDYNNKIDIEFTKSGKIVLNSNCKLKNLVHDIGBEALHTR 470

QY 139 QVFGNNA 146
 DB 471 D-GSSNNS 477

RESULT 13
 US-10-603-150-2
 Sequence 2, Application US/10603150
 GENERAL INFORMATION:
 APPLICANT: Robert G. Wiscotkey
 TITLE OF INVENTION: MAGED3-like Gene Disruptions,
 TITLE OF INVENTION: Compositions and Methods Related Thereto
 FILE REFERENCE: R-1891
 CURRENT APPLICATION NUMBER: US/10/603,150
 PRIOR FILING DATE: 2003-06-24
 PRIOR APPLICATION NUMBER: US 60/391,205
 PRIOR FILING DATE: 2002-06-24
 NUMBER OF SEQ ID NOS: 4
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 2
 LENGTH: 1160
 TYPE: PRT
 ORGANISM: Mus musculus
 US-10-603-150-2

Query Match 9.6%; Score 75; DB 6; Length 1160;
 Best Local Similarity 28.0%; Pred. No. 45;
 Matches 46; Conservative 18; Mismatches 56; Indels 44; Gaps 11;

QY 11 AIVVGSALAGVVPQWCGGNNHGGNSGPDSTL-STIYQSA--NAALYDQVTRVVT 67
 DB 515 AISTFSGVLNSSASFSGAIN-----TSAGFGSTLNSASFSGALSTSFSGVTLN----- 565

QY 68 HEMAHAGYNGADVGGADNSTIELTQNG-----FRNNATIDQNNAKKSDITVQYQY----- 116
 DB 566 ---GRAGFGALNT-----NATFGVLNSAGFGGAMNTVATFG--GALNSNAGFGGALS 615

QY 117 ---QYGG--NNAA---LVNOTASDSVMMQVQFGN--NATAN 149
 DB 616 TSTNFGALNNSAGFGAMTASFGVGLNNSAGFGGAINTSAN 659

RESULT 14
 US-10-771-241-364
 Sequence 364, Application US/10771241
 GENERAL INFORMATION:
 APPLICANT: Zykkind, Judith
 APPLICANT: Feisylh, R. Allyn
 TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
 TITLE OF INVENTION: ESCHERICHIA COLI
 FILE REFERENCE: ELITRA.001CI
 CURRENT APPLICATION NUMBER: US/10/771,241
 PRIOR FILING DATE: 2004-02-03
 PRIOR APPLICATION NUMBER: 09/492,709
 PRIOR FILING DATE: 2000-01-27
 PRIOR APPLICATION NUMBER: 60/117,405
 PRIOR FILING DATE: 1999-01-27
 NUMBER OF SEQ ID NOS: 485
 SOFTWARE: FastSeq for Windows Version 3.0
 SEQ ID NO 364
 LENGTH: 878
 TYPE: PRT
 ORGANISM: E. Coli
 US-10-771-241-364

Query Match 9.4%; Score 73.5; DB 6; Length 878;
 Best Local Similarity 28.7%; Pred. No. 43;

Matches 41; Conservative 18; Mismatches 45; Indels 39; Gaps 10;

```

QY      26 WGGGGHNGGSSGSDTLSTIQ--YGSANAL-----YDQVTRVYTHEMAHAGYGG 78
      :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      655 YAGGGGNGSG--STG-YATLN-YRGYGANIGYSHDDIKOLYYGSGGLAHA--NG 707
      :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY      79 ADVGQ-----GADNSTIELTONGFRNNATIDQNAKNSDITVGQYCGNNALV 126
      :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      708 VTLGGPLNDTVLVKAPGAKDAKE-NQTVRTD-----WRGVAVLPVATEYRENRVALD 761
      :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY      127 NOTASDSSVMRQVGFNNATAN 149
      :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      762 TMTLADNVDL-----DNAVAN 777
      :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

```

RESULT 15

```

US-10-627-556-392
; Sequence 392, Application US/10627556
; GENERAL INFORMATION:
; APPLICANT: LEDBETTER, JEFFREY A.
; APPLICANT: HAYDEN-LEDBETTER, MARTHA
; APPLICANT: THOMPSON, PETER A.
; TITLE OF INVENTION: BINDING CONSTRUCTS AND METHODS FOR USE THEREOF
; FILE REFERENCE: 49076.000004.CIP2
; CURRENT APPLICATION NUMBER: US/10/627,556
; PRIOR FILING DATE: 2003-07-26
; PRIOR APPLICATION NUMBER: 10/053,530
; PRIOR FILING DATE: 2002-01-17
; PRIOR APPLICATION NUMBER: 60/367,358
; PRIOR FILING DATE: 2002-01-16
; PRIOR APPLICATION NUMBER: 09/765,208
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/385,691
; PRIOR FILING DATE: 2002-06-03
; NUMBER OF SEQ ID NOS: 699
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 392
; LENGTH: 273
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: amino acid sequence
US-10-627-556-392

```

Query Match 9.3%; Score 72.5; DB 6; Length 273;

Best Local Similarity 25.8%; Pred. No. 13; Mismatches 36; Indels 15; Gaps 2;

```

QY      15 SCSALAGVPPWGGGHNHNGGSSGSDTLSTIQYGSANALY-----DQVTRVYT 67
      :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      75 ASNLVSGIPIPRFSGS-----SGTDFLTNIHVEKVDATYHCQSTEDPWTFGGCT 126
      :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY      68 HEMAHAGYNGADVCGAGDNSTIELTONG 96
      :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      127 KLEIKGGGSGGSGGSGGSSGVQLQDSG 155
      :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

```

Search completed: March 11, 2004, 19:14:19
Job time : 5.3 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:33:51 ; Search time 171.3 Seconds

(without alignments)
860.386 Million cell updates/sec

Title: US-09-543-407-28

Perfect score: 775
Sequence: 1 MKLLKVAAPAAIVVSGSALA.....DSSVMRVQVFGNNATRAY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 6019581 seqs, 976053577 residues

Total number of hits satisfying chosen parameters: 6019581

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents AA Main.*
1: /cgn2_6/prodata/2/paa/PCRTUS_COMB.pep.*
2: /cgn2_6/prodata/2/paa/US06_COMB.pep.*
3: /cgn2_6/prodata/2/paa/US07_COMB.pep.*
4: /cgn2_6/prodata/2/paa/US08_COMB.pep.*
5: /cgn2_6/prodata/2/paa/US081_COMB.pep.*
6: /cgn2_6/prodata/2/paa/US082_COMB.pep.*
7: /cgn2_6/prodata/2/paa/US083_COMB.pep.*
8: /cgn2_6/prodata/2/paa/US084_COMB.pep.*
9: /cgn2_6/prodata/2/paa/US085_COMB.pep.*
10: /cgn2_6/prodata/2/paa/US086_COMB.pep.*
11: /cgn2_6/prodata/2/paa/US087_COMB.pep.*
12: /cgn2_6/prodata/2/paa/US088_COMB.pep.*
13: /cgn2_6/prodata/2/paa/US089_COMB.pep.*
14: /cgn2_6/prodata/2/paa/US090_COMB.pep.*
15: /cgn2_6/prodata/2/paa/US091_COMB.pep.*
16: /cgn2_6/prodata/2/paa/US092_COMB.pep.*
17: /cgn2_6/prodata/2/paa/US093_COMB.pep.*
18: /cgn2_6/prodata/2/paa/US094_COMB.pep.*
19: /cgn2_6/prodata/2/paa/US095_COMB.pep.*
20: /cgn2_6/prodata/2/paa/US096_COMB.pep.*
21: /cgn2_6/prodata/2/paa/US097A_COMB.pep.*
22: /cgn2_6/prodata/2/paa/US097B_COMB.pep.*
23: /cgn2_6/prodata/2/paa/US098_COMB.pep.*
24: /cgn2_6/prodata/2/paa/US099A_COMB.pep.*
25: /cgn2_6/prodata/2/paa/US099B_COMB.pep.*
26: /cgn2_6/prodata/2/paa/US100_COMB.pep.*
27: /cgn2_6/prodata/2/paa/US101_COMB.pep.*
28: /cgn2_6/prodata/2/paa/US102_COMB.pep.*
29: /cgn2_6/prodata/2/paa/US103_COMB.pep.*
30: /cgn2_6/prodata/2/paa/US104_COMB.pep.*
31: /cgn2_6/prodata/2/paa/US106_COMB.pep.*
32: /cgn2_6/prodata/2/paa/US107_COMB.pep.*
33: /cgn2_6/prodata/2/paa/US107_COMB.pep.*
34: /cgn2_6/prodata/2/paa/US107_COMB.pep.*
35: /cgn2_6/prodata/2/paa/US107_COMB.pep.*
36: /cgn2_6/prodata/2/paa/US107_COMB.pep.*
37: /cgn2_6/prodata/2/paa/US107_COMB.pep.*
38: /cgn2_6/prodata/2/paa/US107_COMB.pep.*
39: /cgn2_6/prodata/2/paa/US107_COMB.pep.*
40: /cgn2_6/prodata/2/paa/US107_COMB.pep.*
41: /cgn2_6/prodata/2/paa/US107_COMB.pep.*
42: /cgn2_6/prodata/2/paa/US107_COMB.pep.*
43: /cgn2_6/prodata/2/paa/US107_COMB.pep.*
44: /cgn2_6/prodata/2/paa/US107_COMB.pep.*
45: /cgn2_6/prodata/2/paa/US107_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
------------	-------	--------------------	----	-------------

1	775	100.0	151	19	US-09-543-407-28	Sequence 28, Appl
2	712	91.9	151	19 <th>US-09-543-407-22</th> <th>Sequence 22, Appl</th>	US-09-543-407-22	Sequence 22, Appl
3	683	88.1	151	19 <th>US-09-543-407-5</th> <th>Sequence 5, Appl</th>	US-09-543-407-5	Sequence 5, Appl
4	678	87.5	151	6 <th>US-08-233-642A-57</th> <th>Sequence 57, Appl</th>	US-08-233-642A-57	Sequence 57, Appl
5	662	85.4	151	19 <th>US-09-543-407-20</th> <th>Sequence 20, Appl</th>	US-09-543-407-20	Sequence 20, Appl
6	617	79.6	151	19 <th>US-09-543-407-26</th> <th>Sequence 26, Appl</th>	US-09-543-407-26	Sequence 26, Appl
7	613	79.1	151	19 <th>US-09-543-407-30</th> <th>Sequence 30, Appl</th>	US-09-543-407-30	Sequence 30, Appl
8	605	78.1	151	19 <th>US-09-543-407-12</th> <th>Sequence 12, Appl</th>	US-09-543-407-12	Sequence 12, Appl
9	603	77.8	151	19 <th>US-09-543-407-14</th> <th>Sequence 14, Appl</th>	US-09-543-407-14	Sequence 14, Appl
10	601	77.5	151	19 <th>US-09-543-407-18</th> <th>Sequence 18, Appl</th>	US-09-543-407-18	Sequence 18, Appl
11	600	77.4	151	19 <th>US-09-543-407-24</th> <th>Sequence 24, Appl</th>	US-09-543-407-24	Sequence 24, Appl
12	596	76.9	151	19 <th>US-09-543-407-31</th> <th>Sequence 31, Appl</th>	US-09-543-407-31	Sequence 31, Appl
13	568	73.3	151	19 <th>US-09-543-407-16</th> <th>Sequence 16, Appl</th>	US-09-543-407-16	Sequence 16, Appl
14	509	65.7	151	19 <th>US-09-543-407-7</th> <th>Sequence 7, Appl</th>	US-09-543-407-7	Sequence 7, Appl
15	506	65.3	151	13 <th>US-08-978-878-4</th> <th>Sequence 4, Appl</th>	US-08-978-878-4	Sequence 4, Appl
16	506	65.3	151	21 <th>US-09-741-873B-4</th> <th>Sequence 2, Appl</th>	US-09-741-873B-4	Sequence 2, Appl
17	504	65.0	151	33 <th>US-60-352-946-2</th> <th>Sequence 2, Appl</th>	US-60-352-946-2	Sequence 2, Appl
18	504	65.0	151	33 <th>US-60-444-371-2</th> <th>Sequence 2, Appl</th>	US-60-444-371-2	Sequence 2, Appl
19	498	64.3	120	6 <th>US-08-233-642A-55</th> <th>Sequence 55, Appl</th>	US-08-233-642A-55	Sequence 55, Appl
20	464	59.9	109	19 <th>US-09-543-407-34</th> <th>Sequence 34, Appl</th>	US-09-543-407-34	Sequence 34, Appl
21	462	59.6	158	16 <th>US-09-252-691-5834</th> <th>Sequence 5834, Ap</th>	US-09-252-691-5834	Sequence 5834, Ap
22	462	59.6	158	16 <th>US-09-252-691C-5834</th> <th>Sequence 5834, Ap</th>	US-09-252-691C-5834	Sequence 5834, Ap
23	462	59.6	158	30 <th>US-10-417-886-5834</th> <th>Sequence 5834, Ap</th>	US-10-417-886-5834	Sequence 5834, Ap
24	428	55.2	131	13 <th>US-08-978-878-2</th> <th>Sequence 2, Appl</th>	US-08-978-878-2	Sequence 2, Appl
25	428	55.2	131	21 <th>US-09-741-873B-2</th> <th>Sequence 2, Appl</th>	US-09-741-873B-2	Sequence 2, Appl
26	324	41.8	109	19 <th>US-09-543-407-35</th> <th>Sequence 35, Appl</th>	US-09-543-407-35	Sequence 35, Appl
27	257	33.2	68	19 <th>US-09-543-407-37</th> <th>Sequence 37, Appl</th>	US-09-543-407-37	Sequence 37, Appl
28	181	23.4	48	19 <th>US-09-543-407-39</th> <th>Sequence 39, Appl</th>	US-09-543-407-39	Sequence 39, Appl
29	160.5	20.7	70	19 <th>US-09-543-407-32</th> <th>Sequence 32, Appl</th>	US-09-543-407-32	Sequence 32, Appl
30	102.5	13.2	186	16 <th>US-09-252-691-5833</th> <th>Sequence 5833, Ap</th>	US-09-252-691-5833	Sequence 5833, Ap
31	102.5	13.2	186	16 <th>US-10-417-886-5833</th> <th>Sequence 5833, Ap</th>	US-10-417-886-5833	Sequence 5833, Ap
32	102.5	13.2	186	30 <th>US-10-417-886-5833</th> <th>Sequence 18734, A</th>	US-10-417-886-5833	Sequence 18734, A
33	95	12.3	400	21 <th>US-09-733-088-18734</th> <th>Sequence 18734, A</th>	US-09-733-088-18734	Sequence 18734, A
34	95	12.3	400	23 <th>US-09-816-660-18734</th> <th>Sequence 18734, A</th>	US-09-816-660-18734	Sequence 18734, A
35	95	12.3	400	27 <th>US-10-155-881-28424</th> <th>Sequence 28424, A</th>	US-10-155-881-28424	Sequence 28424, A
36	95	12.3	400	30 <th>US-10-437-963-186417</th> <th>Sequence 186417, A</th>	US-10-437-963-186417	Sequence 186417, A
37	95	12.3	1060	30 <th>US-10-438-246-117264</th> <th>Sequence 117264, A</th>	US-10-438-246-117264	Sequence 117264, A
38	95	12.3	1060	21 <th>US-09-733-088-18735</th> <th>Sequence 18735, A</th>	US-09-733-088-18735	Sequence 18735, A
39	95	12.3	1060	23 <th>US-09-816-660-18735</th> <th>Sequence 18735, A</th>	US-09-816-660-18735	Sequence 18735, A
40	94.5	12.2	151	19 <th>US-09-543-407-6</th> <th>Sequence 6, Appl</th>	US-09-543-407-6	Sequence 6, Appl
41	92.5	11.9	644	22 <th>US-09-791-537-41783</th> <th>Sequence 41783, A</th>	US-09-791-537-41783	Sequence 41783, A
42	92	11.9	588	22 <th>US-09-791-537-108562</th> <th>Sequence 108562, A</th>	US-09-791-537-108562	Sequence 108562, A
43	92	11.9	1074	28 <th>US-10-282-1222A-50616</th> <th>Sequence 50616, A</th>	US-10-282-1222A-50616	Sequence 50616, A
44	91.5	11.8	478	22 <th>US-09-791-537-22883</th> <th>Sequence 22883, A</th>	US-09-791-537-22883	Sequence 22883, A
45	91.5	11.8	602	22 <th>US-09-791-537-29231</th> <th>Sequence 29231, A</th>	US-09-791-537-29231	Sequence 29231, A

ALIGNMENTS

RESULT 1
US-09-543-407-28
Sequence 28, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543,407
CURRENT FILING DATE: 2000-04-05
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 28
LENGTH: 151
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
OTHER INFORMATION: encoding PT3 from Gpe3 of Leishmania major.

US-09-543-407-28

Query Match 100.0%; Score 775; DB 19; Length 151;
Best Local Similarity 100.0%; Pred. No. 1.2e-74;
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLKVAFPAIIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
DB 1 MKLKVAFPAIIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
QY 61 SDARKSETTTTOSGYGAGADVDOLVTRVYVTHEMAHFRNNATIDQNAKSDITVGOYGG 120
DB 61 SDARKSETTTTOSGYGAGADVDOLVTRVYVTHEMAHFRNNATIDQNAKSDITVGOYGG 120
QY 121 NNAALVNOTASDSSVMVROVGFNNATANOY 151
DB 121 NNAALVNOTASDSSVMVROVGFNNATANOY 151

RESULT 2

US-09-543-407-22

Sequence 22, Application US/09543407

GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 22
LENGTH: 151
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afgA
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from GP63 of *Leishmania major*.
US-09-543-407-22

Query Match 91.9%; Score 712; DB 19; Length 151;
Best Local Similarity 91.1%; Pred. No. 7.2e-68;
Matches 144; Conservative 0; Mismatches 0; Indels 14; Gaps 2;

QY 1 MKLKVAFPAIIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
DB 1 MKLKVAFPAIIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
QY 61 SDARKSETTTTOSGYGAGADVDOLVTRVYVTHEMAHFRNNATIDQNAKSDI 113
DB 61 SDARKSETTTTOSGYGAGADVDOLVTRVYVTHEMAHFRNNATIDQNAKSDI 113
QY 114 TVGOYGNNALVNOTASDSSVMVROVGFNNATANOY 151
DB 114 TVGOYGNNALVNOTASDSSVMVROVGFNNATANOY 151

RESULT 3

US-09-543-407-5

Sequence 5, Application US/09543407

GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407

CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5
LENGTH: 151
TYPE: PRT
ORGANISM: *Salmonella enteritidis*
US-09-543-407-5

Query Match 88.1%; Score 683; DB 19; Length 151;
Best Local Similarity 90.1%; Pred. No. 9.5e-65;
Matches 136; Conservative 1; Mismatches 14; Indels 0; Gaps 0;

QY 1 MKLKVAFPAIIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
DB 1 MKLKVAFPAIIVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60
QY 61 SDARKSETTTTOSGYGAGADVDOLVTRVYVTHEMAHFRNNATIDQNAKSDITVGOYGG 120
DB 61 SDARKSETTTTOSGYGAGADVDOLVTRVYVTHEMAHFRNNATIDQNAKSDITVGOYGG 120
QY 121 NNAALVNOTASDSSVMVROVGFNNATANOY 151
DB 121 NNAALVNOTASDSSVMVROVGFNNATANOY 151

RESULT 4

US-08-233-642A-57

Sequence 57, Application US/08233642A

GENERAL INFORMATION:
APPLICANT: Kay, William W.
APPLICANT: Collinson, S. Karen
APPLICANT: Cloutier, Sharon C.
APPLICANT: Doran, James L.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR *SALMONELLA*-
NUMBER OF SEQUENCES: 58
NUMBER OF SEQUENCES: -
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed and Berry
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: U.S.A.
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233,642A
FILING DATE: 26-Apr-1994
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: King, Joshua
REGISTRATION NUMBER: 35,570
REFERENCE/DOCKET NUMBER: 920043.403C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEDANBERRY
INFORMATION FOR SEQ ID NO: 57:
SEQUENCE CHARACTERISTICS:
LENGTH: 151 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-233-642A-57

Query Match 87.5%; Score 678; DB 6; Length 151;
Best Local Similarity 89.4%; Pred. No. 3.3e-64;
Matches 135; Conservative 1; Mismatches 15; Indels 0; Gaps 0;

```

Qy      1  MKLLVAAFAALIVGSSALAGVVPQWGGGNNHGGNSSGPDSTLSIYQGSANALALQ 60
Db      1  MKLLVAAFAALIVGSSALAGVVPQWGGGNNHGGNSSGPDSTLSIYQGSANALALQ 60
Qy      61  SPARKSETTITSGYNGADVDQLVTRVVTHEMAHAFRNNATIDQNNAKNSDITVQYGG 120
Db      61  SPARKSETTITSGYNGADVDQGGADNSTIELQNFRNNATIIDQNNAKNSDITVQYGG 120
Qy      121  NNALVNOTASDSSVMVRQVGFNNATANOY 151
Db      121  NNPALVNOTASDSSVMVRQVGFNNATANOY 151

RESULT 5
US-09-543-407-20
; Sequence 20, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-20

Query Match      85.4%; Score 662; DB 19; Length 151;
Best Local Similarity 81.9%; Pred. No. 1.7e-62;
Matches 136; Conservative 0; Mismatches 0; Indels 30; Gaps 2

Qy      1  MKLLVAAFAALIVGSSALAGVVPQWGGGNNHGGNSSGPDSTLSIYQGSANALALQ 60
Db      1  MKLLVAAFAALIVGSSALAGVVPQWGGGNNHGGNSSGPDSTLSIYQGSANALALQ 60
Qy      61  SPARKSETTITSGYNGADVDQLVTRVVTHEMAHA-----FNNATIDQ 105
Db      61  SPARK-----YDQLVTRVVTHEMAHAGCGADNSTIELTQNGFNNATIDQ 105
Qy      106  WNAKSDITVGOYGNNALVNOTASDSSVMVRQVGFNNATANOY 151
Db      106  WNAKSDITVGOYGNNALVNOTASDSSVMVRQVGFNNATANOY 151

RESULT 6
US-09-543-407-26
; Sequence 26, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 151

```

```

; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-26

Query Match      79.6%; Score 617; DB 19; Length 151;
Best Local Similarity 73.6%; Pred. No. 1.2e-57;
Matches 128; Conservative 0; Mismatches 0; Indels 46; Gaps 2;

OY      1 MKLTKVAFAFIAIVSGSALAGVPOWGCGGNHNGGSSGPDSLTSLIYQGSANAALALQ 60
          |||
DB       1 MKLTKVAFAFIAIVSGSALAGVPOWGCGGNHNGGSSGPDSLTSLIYQGSANAAL-- 57
          |||

OY      61 SDARKSETTITQSGYGADVDQLVTREVTHMAHA-----F 97
          |||||
DB       58 -----YDQLVTRVTVTHEMAHAYGNGADVGCADNSTIELTONGF 97
          |||||

OY      98 RNNATTIDOMNAKNSDITVGQYGNNALVNQTASDSSVMVRQVFGNNTATNOY 151
          |||
DB       98 RNNATTIDOMNAKNSDITVGQYGNNALVNQTASDSSVMVRQVFGNNTATNOY 151
          |||

RESULT 7
US-09-543-407-30
; Sequence 30, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543.407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-30

Query Match      79.1%; Score 613; DB 19; Length 151;
Best Local Similarity 74.6%; Pred. No. 3.2e-57;
Matches 129; Conservative 0; Mismatches 0; Indels 44; Gaps 2;

OY      1 MKLTKVAFAFIAIVSGSALAGVPOWGCGGNHNGGSSGPDSLTSLIYQGSANAALALQ 60
          |||
DB       1 MKLTKVAFAFIAIVSGSALAGVPOWGCGGNHNGGSSGPDSLTSLIYQGSANAALALQ 60
          |||

OY      61 SDARKSETTITQSGYGADVDQLVTREVTHMAHAFR 98
          |||||
DB       61 SDARKSETTITQSGYGADVDGCADNSTIELTONGRNNATTIDQLVTRVTHMAHA-- 118
          |||||

OY      99 NNATTIDOMNAKNSDITVGQYGNNALVNQTASDSSVMVRQVFGNNTATNOY 151
          |||
DB       119 -----GNNALVNQTASDSSVMVRQVFGNNTATNOY 151
          |||||

RESULT 8
US-09-543-407-12
; Sequence 12, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
```

APPLICANT: COLLINSON, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543,407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12
LENGTH: 151
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
TITLE OF INVENTION: Recombinant Salmonella enteritidis 3b afga
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-12

Query Match 78.1%; Score 605; DB 19; Length 151;
Best Local Similarity 80.1%; Pred. No. 2,3e-56;
Matches 121; Conservative 5; Mismatches 25; Indels 0; Gaps 0;

QY 1 MRLKVAFAFAIVSGSALAGVVPWGGGNNHGGNSGPDSTLSIYQGSANAALALQ 60
DB 1 MRLKVAFAFAIVSGSALAGVVPWGGGNNHGGNSGPDSTLSIYQGSANAALALQ 60
QY 61 SDARKSETTTTOSGNGADVDQVTRVYTHEMAHAFRNATTDOWNAKNSDITVGOYGG 120
DB 61 SDARKSETTTTOSGNGADVDQVTRVYTHEMAHAFRNATTDOWNAKNSDITVGOYGG 120
QY 121 NNALVNOTASDSSVMVROVGFNNATANQY 151
DB 121 NNALVNOTASDSSVMVROVGFNNATANQY 151

RESULT 9

US-09-543-407-14
Sequence 14, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: COLLINSON, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543,407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 14
LENGTH: 151
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
TITLE OF INVENTION: Recombinant Salmonella enteritidis 3b afga
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-14

Query Match 77.8%; Score 603; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 3,8e-56;
Matches 122; Conservative 4; Mismatches 25; Indels 0; Gaps 0;

QY 1 MRLKVAFAFAIVSGSALAGVVPWGGGNNHGGNSGPDSTLSIYQGSANAALALQ 60
DB 1 MRLKVAFAFAIVSGSALAGVVPWGGGNNHGGNSGPDSTLSIYQGSANAALALQ 60
QY 61 SDARKSETTTTOSGNGADVDQVTRVYTHEMAHAFRNATTDOWNAKNSDITVGOYGG 120
DB 61 SDARKSETTTTOSGNGADVDQVTRVYTHEMAHAFRNATTDOWNAKNSDITVGOYGG 120

QY 121 NNALVNOTASDSSVMVROVGFNNATANQY 151
DB 121 LVTRVYTHEMAHAFRNATTDOWNAKNSDITVGOYGG 151

RESULT 10

US-09-543-407-18
Sequence 18, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: COLLINSON, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543,407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 18
LENGTH: 151
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
TITLE OF INVENTION: Recombinant Salmonella enteritidis 3b afga
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-18

Query Match 77.5%; Score 601; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 6,3e-56;
Matches 122; Conservative 4; Mismatches 25; Indels 0; Gaps 0;

QY 1 MRLKVAFAFAIVSGSALAGVVPWGGGNNHGGNSGPDSTLSIYQGSANAALALQ 60
DB 1 MRLKVAFAFAIVSGSALAGVVPWGGGNNHGGNSGPDSTLSIYQGSANAALALQ 60
QY 61 SDARKSETTTTOSGNGADVDQVTRVYTHEMAHAFRNATTDOWNAKNSDITVGOYGG 120
DB 61 SDARKSETTTTOSGNGADVDQVTRVYTHEMAHAFRNATTDOWNAKNSDITVGOYGG 120
QY 121 NNALVNOTASDSSVMVROVGFNNATANQY 151
DB 121 NNALVNOTASDSSVMVROVGFNNATANQY 151

RESULT 11

US-09-543-407-24
Sequence 24, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: COLLINSON, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543,407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 24
LENGTH: 151
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
TITLE OF INVENTION: Recombinant Salmonella enteritidis 3b afga
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-24

Query Match 77.4%; Score 600; DB 19; Length 151;

Best Local Similarity 81.5%; Pred. No. 8,1e-56;
Matches 123; Conservative 3; Mismatches 25; Indels 0; Gaps 0;

QY 1 MKLLKVAAPFAIIVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYOGSANAALAIQ 60
1 MKLLKVAAPFAIIVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYOGSANAALAIQ 60
Db 61 SDARKSETTITGSGYNGADYDQVTRVVTHEMAHAFRNATIDQWNAKNSDITVGOYGG 120
61 SDARKSETTITGSGYNGADYDQVTRVVTHEMAHAFRNATIDQWNAKNSDITVGOYGG 120
QY 121 NNAALVNOTASDSSVWVROVGFNNATANOY 151
121 NNAALVNOTASDSSVWVROVGFNNATANOY 151
Db 121 HEMAHANOTASDSSVWVROVGFNNATANOY 151

RESULT 12
US-09-543-407-31
Sequence 31, Application US/09543407

GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
TITLE OF INVENTION: BACTERIAL FIBRILAL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543,407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 31
LENGTH: 131
TYPE: PRT
ORGANISM: Salmonella enteritidis
US-09-543-407-31

Query Match 76.3%; Score 596; DB 19; Length 131;
Best Local Similarity 88.5%; Pred. No. 1.8e-55;
Matches 116; Conservative 1; Mismatches 14; Indels 0; Gaps 0;

QY 21 GVPWPGGNNHGGNSGPDSTLSIYOGSANAALAIQSDARKSETTITGSGYNGAD 80
1 GVPWPGGNNHGGNSGPDSTLSIYOGSANAALAIQSDARKSETTITGSGYNGAD 60
Db 81 YDQVTRVVTHEMAHAFRNATIDQWNAKNSDITVGOYGGNNAALVNOTASDSSVWVROV 140
61 VGGADNSTIELTQNGFRNATIDQWNAKNSDITVGOYGGNNAALVNOTASDSSVWVROV 120
QY 141 GFGNNATANOY 151
141 GFGNNATANOY 151
Db 121 GFGNNATANOY 131

RESULT 13
US-09-543-407-16
Sequence 16, Application US/09543407

GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
TITLE OF INVENTION: BACTERIAL FIBRILAL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543,407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 16
LENGTH: 151
TYPE: PRT
ORGANISM: Artificial Sequence

FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
OTHER INFORMATION: Sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-16

Query Match 73.3%; Score 568; DB 19; Length 151;
Best Local Similarity 80.1%; Pred. No. 2.2e-52;
Matches 121; Conservative 4; Mismatches 26; Indels 0; Gaps 0;

QY 1 MKLLKVAAPFAIIVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYOGSANAALAIQ 60
1 MKLLKVAAPFAIIVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYOGSANAALAIQ 60
Db 61 SDARKSETTITGSGYNGADYDQVTRVVTHEMAHAFRNATIDQWNAKNSDITVGOYGG 120
61 SDARKSETTITGSGYNGADYDQVTRVVTHEMAHAFRNATIDQWNAKNSDITVGOYGG 120
QY 121 NNAALVNOTASDSSVWVROVGFNNATANOY 151
121 NNAALVNOTASDSSVWVROVGFNNATANOY 151
Db 121 NNAALVNOTASDSSVWVROVGFNNATANOY 151

RESULT 14
US-09-543-407-7
Sequence 7, Application US/09543407

GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
TITLE OF INVENTION: BACTERIAL FIBRILAL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543,407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 7
LENGTH: 151
TYPE: PRT
ORGANISM: Escherichia coli
US-09-543-407-7

Query Match 65.7%; Score 509; DB 19; Length 151;
Best Local Similarity 68.2%; Pred. No. 5e-46;
Matches 103; Conservative 15; Mismatches 33; Indels 0; Gaps 0;

QY 1 MKLLKVAAPFAIIVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYOGSANAALAIQ 60
1 MKLLKVAAPFAIIVSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYOGSANAALAIQ 60
Db 61 SDARKSETTITGSGYNGADYDQVTRVVTHEMAHAFRNATIDQWNAKNSDITVGOYGG 120
61 SDARKSETTITGSGYNGADYDQVTRVVTHEMAHAFRNATIDQWNAKNSDITVGOYGG 120
QY 121 NNAALVNOTASDSSVWVROVGFNNATANOY 151
121 NNAALVNOTASDSSVWVROVGFNNATANOY 151
Db 121 NNAALVNOTASDSSVWVROVGFNNATANOY 151

RESULT 15
US-08-978-878-4
Sequence 4, Application US/08978878

GENERAL INFORMATION:
APPLICANT: NORMARK, Staffan
APPLICANT: OLSEN, Arne
TITLE OF INVENTION: FIBRONECTIN BINDING PROTEIN AS WELL AS ITS PREPARATION
FILE REFERENCE: 012889-081
CURRENT APPLICATION NUMBER: US/08/978,878
CURRENT FILING DATE: 1997-11-26
EARLIER APPLICATION NUMBER: SE 8801723-1
EARLIER FILING DATE: 1988-05-06

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:35:05 ; Search time 5.3 Seconds

(Without alignments)
376.014 Million cell updates/sec

Title: US-09-543-407-28

Perfect score: 775
Sequence: 1 MKLLKVAAPAAIVSGSALA.....DSSVMRVGVFGNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 125546 seqs, 13197846 residues

Total number of hits satisfying chosen parameters: 125546

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents AA New:
1: /cgn2_6/prodata/1/paa/PCT_NEW_COMB.pep.*
2: /cgn2_6/prodata/1/paa/US06_NEW_COMB.pep.*
3: /cgn2_6/prodata/1/paa/US07_NEW_COMB.pep.*
4: /cgn2_6/prodata/1/paa/US08_NEW_COMB.pep.*
5: /cgn2_6/prodata/1/paa/US09_NEW_COMB.pep.*
6: /cgn2_6/prodata/1/paa/US10_NEW_COMB.pep.*
7: /cgn2_6/prodata/1/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	506	65.3	151	US-09-741-873C-4	Sequence 4, Appl1
2	428	55.2	131	US-09-741-873C-2	Sequence 2, Appl1
3	84.5	10.9	234	US-10-767-701-45603	Sequence 45603, A
4	77	9.9	1160	US-10-603-150-2	Sequence 2, Appl1
5	75.5	9.7	424	US-10-045-674A-591	Sequence 591, App
6	75.5	9.7	533	US-10-045-674A-527	Sequence 527, App
7	74	9.5	447	US-10-767-701-45219	Sequence 45219, A
8	73.5	9.5	382	US-10-771-241-299	Sequence 299, App
9	73	9.4	162	US-10-620-246-6	Sequence 6, Appl1
10	73	9.4	174	US-10-767-701-56903	Sequence 56903, A
11	73	9.4	238	US-10-779-461-1	Sequence 1, Appl1
12	72.5	9.4	240	US-09-596-774-2	Sequence 2, Appl1
13	72.5	9.4	434	US-10-045-674A-594	Sequence 594, App
14	72.5	9.4	443	US-09-596-774-7	Sequence 7, Appl1
15	72.5	9.4	431	US-09-596-774-6	Sequence 6, Appl1
16	72	9.3	461	US-10-767-701-43958	Sequence 43958, A
17	72	9.3	360	US-10-695-499-178	Sequence 178, App
18	72	9.3	628	US-10-695-499-183	Sequence 183, App
19	72	9.3	895	US-10-286-085A-19	Sequence 19, Appl
20	72	9.3	895	US-10-286-085A-20	Sequence 20, Appl
21	72	9.3	895	US-10-286-085A-27	Sequence 27, Appl
22	72	9.3	896	US-10-286-085A-17	Sequence 17, Appl
23	72	9.3	896	US-10-286-085A-18	Sequence 18, Appl
24	72	9.3	896	US-10-286-085A-26	Sequence 26, Appl
25	72	9.3	899	US-10-286-085A-28	Sequence 28, Appl
26	72	9.3	1327	PCT-US04-02338-49	Sequence 49, Appl

27	72	9.3	1532	6	US-10-695-499-62	Sequence 62, Appl
28	71.5	9.2	463	6	US-10-641-678-71	Sequence 71, Appl
29	71.5	9.2	878	6	US-10-771-241-364	Sequence 364, App
30	71	9.2	179	6	US-10-767-701-42696	Sequence 42696, A
31	71	9.2	250	6	US-10-779-461-6	Sequence 6, Appl1
32	71	9.2	514	5	US-09-830-230A-430	Sequence 430, App
33	71	9.2	541	5	US-09-830-230A-429	Sequence 429, App
34	70	9.0	194	6	US-10-767-701-43391	Sequence 43391, A
35	70	9.0	245	6	US-10-779-461-59	Sequence 59, Appl
36	70	9.0	246	6	US-10-779-461-41	Sequence 41, Appl
37	70	9.0	321	1	PCT-US04-05654-590	Sequence 590, App
38	69.5	9.0	310	1	PCT-US04-05654-1593	Sequence 1593, App
39	69.5	9.0	577	6	US-10-775-337-3	Sequence 3, Appl1
40	69.5	9.0	624	1	PCT-US04-05654-1466	Sequence 1466, App
41	69	8.9	182	6	US-10-767-701-44940	Sequence 44940, A
42	69	8.9	351	1	PCT-US04-05654-1755	Sequence 1755, App
43	69	8.9	414	6	US-10-767-701-45857	Sequence 45857, A
44	69	8.9	904	6	US-10-786-892-158	Sequence 158, App
45	69	8.9	1574	6	US-10-695-499-179	Sequence 179, App

ALIGNMENTS

```

RESULT 1
US-09-741-873C-4
; Sequence 4, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsepp, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
; US-09-741-873C-4

```

Query Match 65.3%; Score 506; DB 5; Length 151;

Best local Similarity 67.5%; Pred. No. 6e-39;

Matches 102; Conservative 16; Mismatches 33; Indels 0; Gaps 0;

```

QY 1 MKLLKVAAPAAIVSGSALAGVVPWGCGGNNHNGSSGPDSTLSIYOGSANAALAIQ 60
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1 MKLLKVAAPAAIVSGSALAGVVPWGCGGNNHNGSSGPDSTLSIYOGGNSALAIQ 60
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 61 SPARKETTTTQSGVGNAGADYDLVTRVYTHMAHFRNNATIDONNANRSDITVOYXG 120
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 61 TDRNSDLTITOHGGNGADVGGSDSSIDLTFRGFNSATIDONNANRSEITVAFQFG 120
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 121 NNAALVNOTASDSVWVRGVFGNNATANQY 151
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 121 GNGAAVDOTASNSVAVTVGVFGNNATANQY 151
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

```

RESULT 2
US-09-741-873C-2
Sequence 2, Application US/09741873C
GENERAL INFORMATION:
APPLICANT: Normark, Staffan
APPLICANT: Olsen, Arne
TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
FILE REFERENCE: 012889-084
CURRENT APPEAL NUMBER: US/09/741,873C
CURRENT FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: SE 8601723-1
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: US 08/978,878
PRIOR FILING DATE: 1997-11-26
PRIOR APPLICATION NUMBER: US 07/347,189
PRIOR FILING DATE: 1989-05-04
PRIOR APPLICATION NUMBER: US 07/789,437
PRIOR FILING DATE: 1991-11-06
PRIOR APPLICATION NUMBER: US 07/970,846
PRIOR FILING DATE: 1992-11-03
PRIOR APPLICATION NUMBER: US 08/187,865
PRIOR FILING DATE: 1994-01-28
PRIOR APPLICATION NUMBER: US 08/318,519
PRIOR FILING DATE: 1994-10-05
NUMBER OF SEQ ID NOS: 11
SOFTWARE: PatentIn version 3.0
SEQ ID NO 2
LENGTH: 131
TYPE: prt
ORGANISM: Escherichia coli
US-09-741-873C-2

```

```

Query Match          55.2%; Score 428; DB 5; Length 131;
Best Local Similarity 64.1%; Pred No. 5.3e-32;
Matches      84; Conservative 15; Mismatches 32; Indels 0; Gaps 0;

Oy      21 GVPQWGGGNNHGGNGNSGPDSTLSIYQYGSANALALQSDARKSETTITQSGYNGAD 80
         1 GVPQYGGGGHGGNGGNSGPNSELNTYQYGGGNSALALQTDARNSDLTITQGGGNGAD 60
Oy      81 YDQLYRVYVTHMAHAFRNATITIDQNAKNSDITVQYQYGGNNAALVNOTASDSVWVRQY 140
         61 VQGGSDSSIDLTQRFQGSATLTDQNGKNSLNTVYQFGGGAADVQTA SNSVAVTVQY 120
Oy      141 GFGNNATANQY 151
         |||||:|
Db      121 GFGNNATAHQY 131

RESULT 3
US-10-767-701-45603
; Sequence 45603, Application US/10767701
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(5353)B
; CURRENT APPLICATION NUMBER: US/10/767, 701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 45603
; LENGTH: 234
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(234)
; OTHER INFORMATION: Plants and Uses Thereof For Plant Improvement
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAHY03-C82834_1.pep
; US-10-767-701-45603

```

Query Match	10.9%;	Score 84.5;	DB 6;	Length 234;
Best Local Similarity	24.2%;	Pred. No. 0.93;		
Matches 39;	Conservative 11;	Mismatches 46;	Indels 65;	Gaps 6;

```

QY 1 MLLTKVAAFAALV--SSALAGVPEQW---GGGNNHGGN-----SSGPDSTL 45
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1 MATTKLAALCFIVLLIGGANAAVARVARSAGGGGGGGGGGRGMRGASRFGSGSSGC 60
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 46 SIY-----OYSAFAALALQSDARKSETTITSGYNGADYD 82
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 61 GXYGEAGSGGKAYAOGGGGGGGGGGGGYGS--SSG-----SGYSSGSGYG 102
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 83 QLVTRVVTHEMAFAFNNAITIDQNNAKKSDITVGOYGGNNA 123
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 103 Q-----AGSGSGNGAYAOGGAOOGGGGGGGGGYGSSS 134
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

```

RESULT 4
US-10-603-150-2
; Sequence 2, Application US/10603150
; GENERAL INFORMATION:
; APPLICANT: Robert G. Wisotzkey
; TITLE OF INVENTION: MAGEB3-like Gene Disruptions,
; TITLE OF INVENTION: Compositions and Methods Related Thereof
; FILE REFERENCE: R-1891
; CURRENT APPLICATION NUMBER: US/10/603,150
; PRIORITY FILING DATE: 2003-06-24
; PRIOR APPLICATION NUMBER: US 60/391,205
; PRIOR FILING DATE: 2002-06-24
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1160
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-603-150-2

```

	Query Match	9.9%; Score 77;	DB 6;	Length 1160;
	Best Local Similarity	27.6%;	Pred. No. 32;	
	Matches 43;	Conservative 19;	Mismatches 66;	Indels 28; Gaps 8;
QY	11 AIVVSGSALGAVVPWQMGGNGNNGNSGCEPDTSLIYYQSANAALALOGDARKSETTI	70		
Dd	515 AIISFSGVLNSSASFCGAIN-----TSAGRGSTLN-----SSAFGSALSTSFGVILN	565		
QY	71 TQSGYGNADYDQLVTYRVTHE--MAAIFRNATIIDMNAKNSITVG-----QYGC	120		
Dd	566 GRAEFGALNTNAFFGVILNGSAGFGGAMTNTATFG--GAINSAAGFGAISTSTNFEGA	623		
QY	121 -NNAA-----LVNOTAPDSVVVRQVGFN--NATAN	149		
Dd	624 LNNAGFGGANNTISASFQGVILNNSAGFGGAILNTSAN	659		

```

RESULT 5
US-10-045-674A-591
: Sequence 591, Application US/10045674A
:
: GENERAL INFORMATION:
:
: APPLICANT: LADNER, ROBERT C.
: APPLICANT: COHEN, EDWARD H.
: APPLICANT: NASTRI, HORACIO G.
: APPLICANT: ROOKEY, KRISTIN L.
: APPLICANT: HOET, RENE
:
: APPLICANT: HOOGENBOOM, HENDRICUS R. J. M.
:
: TITLE OF INVENTION: NOVEL METHODS OF CONSTRUCTING LIBRARIES COMPRISING
:
: TITLE OF INVENTION: DISPLAYED AND/OR EXPRESSED MEMBERS OF A DIVERSE FAMILY
:
: TITLE OF INVENTION: OF PEPTIDES, POLYPEPTIDES OR PROTEINS AND THE NOVEL
:
: TITLE OF INVENTION: LIBRARIES
:
: FILE REFERENCE: DYAX/002 CIP2
:
: CURRENT APPLICATION NUMBER: US/10/045,674A
:
: CURRENT FILING DATE: 2001-10-25
:
: PRIORITY APPLICATION NUMBER: 06/198,069
:

```



```

; PRIOR FILING DATE: 2000-04-17
; PRIOR APPLICATION NUMBER: 09/837,306
; PRIOR FILING DATE: 2001-04-17
; NUMBER OF SEQ ID NOS: 635
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 591
; LENGTH: 424
; TYPE: PRT
; ORGANISM: Unknown Organism
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: M13 protein
US-10-045-674A-591

```

```

Query Match          9.7%; Score 75.5; DB 6; Length 424;
Best Local Similarity 25.7%; Pred. No. 13;
Matches 37; Conservative 15; Mismatches 47; Indels 45; Gaps 7;

```

```

QY 15 SGSLAGVVPWGG--GGNHNGGNSGPDSTLSIYQ-YGSANMAL-----ALQSDA 63
DB 243 SGGSEGGGSEGGSEGGSEGGSGDGFDEKXANANKGAMTENADENALQSDA 302
QY 64 R-KSETTITQSGYGNGADYDQLVTRVYTHEMAHAFRNATIDQWAKNSDITVGOYGGNN 122
DB 303 KKKLDSVAT-----DYGAIDGFIQVSGLANGANGAT-----GDPAGSN 341
QY 123 MALVNOTASDSSVMVROVGFGNNA 146
DB 342 SQMA-----QVGDGDN 353

```

RESULT 6
US-10-045-674A-527
Sequence 527, Application US/10045674A

```

; GENERAL INFORMATION:
; APPLICANT: LADNER, ROBERT C.
; APPLICANT: COHEN, EDWARD H.
; APPLICANT: MASTRI, HORACIO G.
; APPLICANT: ROOKEY, KRISTIN L.
; APPLICANT: HOET, RENE
; APPLICANT: HOEGENBOOM, HENDRICUS R. J. M.
; TITLE OF INVENTION: NOVEL METHODS OF CONSTRUCTING LIBRARIES COMPRISING
; TITLE OF INVENTION: DISPLAYED AND/OR EXPRESSED MEMBERS OF A DIVERSE FAMILY
; TITLE OF INVENTION: OF PEPTIDES, POLYPEPTIDES OR PROTEINS AND THE NOVEL
; FILE REFERENCE: DYAX/002 CIP2
; CURRENT APPLICATION NUMBER: US/10/045,674A
; PRIOR FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: 06/198,069
; PRIOR FILING DATE: 2000-04-17
; PRIOR APPLICATION NUMBER: 09/837,306
; PRIOR FILING DATE: 2001-04-17
; NUMBER OF SEQ ID NOS: 635
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 527
; LENGTH: 533
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Vector pCRESS
US-10-045-674A-527

```

```

Query Match          9.7%; Score 75.5; DB 6; Length 533;
Best Local Similarity 25.7%; Pred. No. 17;
Matches 37; Conservative 15; Mismatches 47; Indels 45; Gaps 7;

```

```

QY 15 SGSLAGVVPWGG--GGNHNGGNSGPDSTLSIYQ-YGSANMAL-----ALQSDA 63
DB 352 SGGSEGGGSEGGSEGGSEGGSGDGFDEKXANANKGAMTENADENALQSDA 411
QY 64 R-KSETTITQSGYGNGADYDQLVTRVYTHEMAHAFRNATIDQWAKNSDITVGOYGGNN 122

```

```

DB 412 KKKLDSVAT-----DYGAIDGFIQVSGLANGANGAT-----GDPAGSN 450
QY 123 MALVNOTASDSSVMVROVGFGNNA 146
DB 451 SQMA-----QVGDGDN 462

```

RESULT 7
US-10-767-701-45219
Sequence 45219, Application US/10767701

```

; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(5353)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; PRIOR FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 45219
; LENGTH: 447
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAV03-C9039_1.pep
US-10-767-701-45219

```

```

Query Match          9.5%; Score 74; DB 6; Length 447;
Best Local Similarity 25.3%; Pred. No. 18;
Matches 40; Conservative 25; Mismatches 51; Indels 42; Gaps 9;

```

```

QY 24 PWWGGGNNHNGGNSGPDSTLSIYQGSANMALALQSDARKSETTITQSGYGN- 77
DB 8 PSYGYGYDYDQVSGTSSGNSNTGNTSGY---SARYAPSSBNVQPTTARLQKYSRI 64
QY 78 GADYDQLVTRVYTHEMAHAFRNATIDQWAKNSDITVGOYGGNN 121
DB 65 GDDYRSL--SQVTEALQAGLESNLIYQIDFTKSNEMWTKISFNRCLHDI-----GN 116
QY 122 N-----MALVNOTAS--DSSVMVROVGFGNNA 150
DB 117 TPNPEQAISITIGRLSAFDEDNLI PCFGFDASTHDQ 154

```

RESULT 8
US-10-771-241-299

```

; Sequence 299, Application US/10771241
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Allen
; APPLICANT: Zyskind, Judith
; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
; TITLE OF INVENTION: ESCHERICHIA COLI
; FILE REFERENCE: ELITRA.001C1
; CURRENT APPLICATION NUMBER: US/10/771,241
; PRIOR FILING DATE: 2004-02-03
; PRIOR APPLICATION NUMBER: 09/492,709
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: 60/117,405
; PRIOR FILING DATE: 1999-01-27
; NUMBER OF SEQ ID NOS: 485
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 299
; LENGTH: 382
; TYPE: PRT
; ORGANISM: E. Coli
US-10-771-241-299

```

```

Query Match          9.5%; Score 73.5; DB 6; Length 382;
Best Local Similarity 25.2%; Pred. No. 17;
Matches 29; Conservative 11; Mismatches 40; Indels 35; Gaps 4;

```

```

QY 30 GNNHNGGNSGPDSTLSI-YQ-YGSANMALALQSDARKSETTITQSGYGNGADYDQLV 86

```

Db 158 GHTHGATNTSSGYSGLNRYKXNTNV-----GYSRGDSQIY 199
 QY 87 RVVTHEMAHAFNNATIDQMAKNSDITVGYGNNALVNOTASDSSVMYQV 141
 200 GMSGGI IAH-----DGIITGQPLGDMVLVKAPGADNVKLENQTG 240

RESULT 9
 US-10-620-246-6
 ; Sequence 6, Application US/10620246
 ; GENERAL INFORMATION:

APPLICANT: ANDERSEN, Peter
 APPLICANT: NIELSEN, Rikke
 APPLICANT: OETTINGER, Thomas
 APPLICANT: RASMUSSEN, Peter Birk
 APPLICANT: ROSENKRANDS, Ida
 APPLICANT: WELDINGH, Karin
 APPLICANT: FLORIO, Walter
 TITLE OF INVENTION: NUCLEIC ACIDS FRAGMENTS AND POLYPEPTIDE FRAGMENTS
 TITLE OF INVENTION: DERIVED FROM M. TUBERCULOSIS
 FILE REFERENCE: 670001-2002.1A
 CURRENT APPLICATION NUMBER: US/10/620,246
 CURRENT FILING DATE: 2003-07-15

PRIOR APPLICATION NUMBER: 09/050,739
 PRIOR FILING DATE: 1998-03-30
 PRIOR APPLICATION NUMBER: 0376/97
 PRIOR FILING DATE: 1997-04-02
 PRIOR APPLICATION NUMBER: 1277/97
 PRIOR FILING DATE: 1997-11-10
 PRIOR APPLICATION NUMBER: 60/044,624
 PRIOR FILING DATE: 1997-04-18
 PRIOR APPLICATION NUMBER: 60/070,488
 PRIOR FILING DATE: 1998-01-05
 PRIOR APPLICATION NUMBER: 10/138,473
 PRIOR FILING DATE: 2002-05-02
 PRIOR APPLICATION NUMBER: 09/791,171
 PRIOR FILING DATE: 2001-02-20
 PRIOR APPLICATION NUMBER: 09/415,884
 PRIOR FILING DATE: 1999-10-08
 PRIOR APPLICATION NUMBER: 60/116,673
 PRIOR FILING DATE: 1999-01-21
 PRIOR APPLICATION NUMBER: 1281/98
 PRIOR FILING DATE: 1998-10-08
 NUMBER OF SEQ ID NOS: 173
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 6
 LENGTH: 162
 TYPE: PRT
 ORGANISM: Mycobacterium tuberculosis
 US-10-620-246-6

Query Match 9.4%; Score 73; DB 6; Length 162;
 Best Local Similarity 23.3%; Pred. No. 6.4;
 Matches 24; Conservative 22; Mismatches 47; Indels 10; Gaps 3;
 QY 49 QYGANALALQSDARKSETTITQSGYNGADY--DQVTRVVTHEMAHAFNNATIDQW 106
 Db 38 QAGTESAVSGVGLPPGSLLVKRGPNAGSRFLDQATISAGRHPSDIFLDVTVRR 97
 QY 107 NA---KNSDITVGYGNNALVNOTASDSSVMYR---QVG 141
 Db 98 HAEFLNENFNVVDVGLNGTVYVREPVDSAVLANGEVQIG 140

RESULT 10
 US-10-767-701-56903
 ; Sequence 56903, Application US/10767701
 ; GENERAL INFORMATION:
 APPLICANT: KOVALLIC, David K.
 APPLICANT: Zhou, Yihua
 APPLICANT: Cao, Yongwei
 TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
 FILE REFERENCE: 38-21(5353)B
 CURRENT APPLICATION NUMBER: US/10/767,701
 CURRENT FILING DATE: 2004-01-29
 NUMBER OF SEQ ID NOS: 63128
 SEQ ID NO 56903
 LENGTH: 174
 TYPE: PRT
 ORGANISM: Sorghum bicolor
 FEATURE:
 OTHER INFORMATION: Clone ID: 30950641.pcp
 US-10-767-701-56903

Query Match 9.4%; Score 73; DB 6; Length 174;
 Best Local Similarity 24.5%; Pred. No. 7;
 Matches 35; Conservative 22; Mismatches 54; Indels 32; Gaps 6;

QY 3 LKVAFAIIVSGSALAGVPPQGGGNNHG--GNSGSDSTLSIYQGSANALALQ 60
 Db 6 MIAAFAVAVLAPPAASSGPTPAPALGINYQVADNLPPQAMLLRLALNATRVKLYD 65
 QY 61 SPARKSETTITQSGYNGADY--DQVTRVVTHEMAHAFNNATIDQNNAKY----- 110
 Db 66 ADAR-----VLRAFGSGVDFTVGVPDRLVPRMATDPSAAA-----SWRSNLLPRL 112
 QY 111 -----SDITVQ--YGGNNALV 126
 Db 113 PATSITAVTVGNEVLSTGTAAML 135

RESULT 11
 US-10-779-461-1
 ; Sequence 1, Application US/10779461
 ; GENERAL INFORMATION:
 APPLICANT: MORTON, Philip A
 TITLE OF INVENTION: ANTIBODIES TO C-MET FOR THE TREATMENT OF CANCERS
 FILE REFERENCE: 00980/1
 CURRENT APPLICATION NUMBER: US/10/779,461
 CURRENT FILING DATE: 2004-02-13
 PRIOR APPLICATION NUMBER: 60/447,073
 PRIOR FILING DATE: 2003-02-13
 NUMBER OF SEQ ID NOS: 161
 SOFTWARE: PatentIn version 3.2
 SEQ ID NO 1
 LENGTH: 238
 TYPE: PRT
 ORGANISM: artificial
 FEATURE:
 OTHER INFORMATION: phage display generated human antibody
 US-10-779-461-1

Query Match 9.4%; Score 73; DB 6; Length 238;
 Best Local Similarity 33.3%; Pred. No. 10;
 Matches 24; Conservative 5; Mismatches 19; Indels 24; Gaps 4;
 QY 25 QMG-----GGNNGGNSGPDSTLSIYQGSANALALQSD---ARKSETTITQ 72
 Db 98 RMGGTTVTVSSGGGSGGSGG-----GSAQVLTQPSVSAQPGQRTVISC 147
 QY 73 SGYGN--GADYD 82
 Db 148 TGSSNIGADYD 159

RESULT 12
 US-09-596-774-2
 ; Sequence 2, Application US/09596774
 ; GENERAL INFORMATION:
 APPLICANT: GRONER, BERND
 APPLICANT: MORITZ, DIRK
 TITLE OF INVENTION: BIFUNCTIONAL PROTEIN, PREPARATION AND USE
 FILE REFERENCE: 24741-1521
 CURRENT APPLICATION NUMBER: US/09/596,774

/ CURRENT FILING DATE: 2000-06-19
 / PRIOR APPLICATION NUMBER: 09/159,027
 / PRIOR FILING DATE: 1998-09-23
 / PRIOR APPLICATION NUMBER: 08/793,048
 / PRIOR FILING DATE: 1996-11-01
 / PRIOR APPLICATION NUMBER: PCT/EP95/01494
 / PRIOR FILING DATE: 1995-04-20
 / PRIOR APPLICATION NUMBER: EP 94810244.7
 / NUMBER OF SEQ ID NOS: 16
 / SOFTWARE: PatentIn Ver. 3.2
 / SEQ ID NO 2
 / LENGTH: 240
 / TYPE: PRF
 / ORGANISM: Artificial Sequence
 / FEATURE:
 / OTHER INFORMATION: Description of Artificial Sequence: Synthetic
 / US-09-596-774-2

Query Match 9.4%; Score 72.5; DB 5; Length 240;
 Best Local Similarity 21.2%; Pred. No. 12;
 Matches 25; Conservative 14; Mismatches 38; Indels 41; Gaps 3;

QY 21 GVPQMG-----GGNNHGGNSGPDSTLSI-----47
 DB 104 GVPYWGQGTTVTVSSGGGSGGSDIQLQSHKFLSTVGDVRSITCKASQDV 163

QY 48 -----YQY--GSANAALALQSDARKSETTITQSGYNGADYQDLVTRVTHEMAAF 97
 DB 164 YNAVAMYQKQKQSPKLLIYSASRHYTGVPSTRFTSGSGPDFTFTISSVOAEDLAVYF 221

RESULT 13
 / US-10-045-674A-594
 / Sequence 594, Application US/10045674A
 / GENERAL INFORMATION:
 / APPLICANT: LADNER, ROBERT C.
 / APPLICANT: COHEN, EDWARD H.
 / APPLICANT: NASTRI, HORACIO G.
 / APPLICANT: ROOKEY, KRISTIN L.
 / APPLICANT: HOET, RENE
 / APPLICANT: HOOGENBOOM, HENDRICUS R. J. M.
 / TITLE OF INVENTION: NOVEL METHODS OF CONSTRUCTING LIBRARIES COMPRISING
 / TITLE OF INVENTION: DISPLAYED AND/OR EXPRESSED MEMBERS OF A DIVERSE FAMILY
 / TITLE OF INVENTION: OF PEPTIDES, POLYPEPTIDES OR PROTEINS AND THE NOVEL
 / FILE REFERENCE: DYA/002 CIP2
 / CURRENT APPLICATION NUMBER: US/10/045,674A
 / CURRENT FILING DATE: 2001-10-25
 / PRIOR APPLICATION NUMBER: 06/198,069
 / PRIOR FILING DATE: 2000-04-17
 / PRIOR APPLICATION NUMBER: 09/837,306
 / PRIOR FILING DATE: 2001-04-17
 / NUMBER OF SEQ ID NOS: 635
 / SOFTWARE: PatentIn Ver. 2.1
 / SEQ ID NO 594
 / LENGTH: 434
 / TYPE: PRF
 / ORGANISM: Artificial Sequence
 / FEATURE:
 / OTHER INFORMATION: Description of Artificial Sequence: M13-III
 / US-10-045-674A-594

Query Match 9.4%; Score 72.5; DB 6; Length 434;
 Best Local Similarity 25.6%; Pred. No. 24;
 Matches 34; Conservative 14; Mismatches 38; Indels 47; Gaps 7;

QY 28 GGGNNGG--GNSGPDSTLSIYQY--GSANAAL-----ALQSDAR--SETTITQS 73
 DB 264 GGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSDIQLQSHKFLSTVGDVRSITCKASQDV 321

QY 74 GYNGADYQDLVTRVTHEMAAFERNNATIDQNAKNSDITVGOYGGNNAALVQRTASDS 133
 DB 322 -----DYGAIDIGFIDVSGGLANGCAT-----GDFAGNSGMA-----355
 QY 134 SWVRQVFGNNA 146
 DB 356 -----QVGDGDSN 363

RESULT 14
 / US-09-596-774-7
 / Sequence 7, Application US/09596774
 / GENERAL INFORMATION:
 / APPLICANT: GRONER, BERND
 / APPLICANT: MORITZ, DIRK
 / TITLE OF INVENTION: BIFUNCTIONAL PROTEIN, PREPARATION AND USE
 / FILE REFERENCE: 24741-1521
 / CURRENT APPLICATION NUMBER: US/09/596,774
 / CURRENT FILING DATE: 2000-06-19
 / PRIOR APPLICATION NUMBER: 09/159,027
 / PRIOR FILING DATE: 1998-09-23
 / PRIOR APPLICATION NUMBER: 08/793,048
 / PRIOR FILING DATE: 1996-11-01
 / PRIOR APPLICATION NUMBER: PCT/EP95/01494
 / PRIOR FILING DATE: 1995-04-20
 / PRIOR APPLICATION NUMBER: EP 94810244.7
 / PRIOR FILING DATE: 1994-05-02
 / NUMBER OF SEQ ID NOS: 16
 / SOFTWARE: PatentIn Ver. 3.2
 / SEQ ID NO 7
 / LENGTH: 443
 / TYPE: PRF
 / ORGANISM: Artificial Sequence
 / FEATURE:
 / OTHER INFORMATION: Description of Artificial Sequence: Synthetic
 / US-09-596-774-7

Query Match 9.4%; Score 72.5; DB 5; Length 443;
 Best Local Similarity 21.2%; Pred. No. 25;
 Matches 25; Conservative 14; Mismatches 38; Indels 41; Gaps 3;

QY 21 GVPQMG-----GGNNHGGNSGPDSTLSI-----47
 DB 104 GVPYWGQGTTVTVSSGGGSGGSDIQLQSHKFLSTVGDVRSITCKASQDV 163

QY 48 -----YQY--GSANAALALQSDARKSETTITQSGYNGADYQDLVTRVTHEMAAF 97
 DB 164 YNAVAMYQKQKQSPKLLIYSASRHYTGVPSTRFTSGSGPDFTFTISSVOAEDLAVYF 221

RESULT 15
 / US-09-596-774-6
 / Sequence 6, Application US/09596774
 / GENERAL INFORMATION:
 / APPLICANT: GRONER, BERND
 / APPLICANT: MORITZ, DIRK
 / TITLE OF INVENTION: BIFUNCTIONAL PROTEIN, PREPARATION AND USE
 / FILE REFERENCE: 24741-1521
 / CURRENT APPLICATION NUMBER: US/09/596,774
 / CURRENT FILING DATE: 2000-06-19
 / PRIOR APPLICATION NUMBER: 09/159,027
 / PRIOR FILING DATE: 1998-09-23
 / PRIOR APPLICATION NUMBER: 08/793,048
 / PRIOR FILING DATE: 1996-11-01
 / PRIOR APPLICATION NUMBER: PCT/EP95/01494
 / PRIOR FILING DATE: 1995-04-20
 / PRIOR APPLICATION NUMBER: EP 94810244.7
 / PRIOR FILING DATE: 1994-05-02
 / NUMBER OF SEQ ID NOS: 16
 / SOFTWARE: PatentIn Ver. 3.2
 / SEQ ID NO 6
 / LENGTH: 461

Fri Mar 12 09:42:04 2004

us-09-543-407-28.rapn

```

; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: chimeric amino acid sequence
; US-09-536-774-6

Query Match          9.4%; Score 72.5; DB 5; Length 461;
Best Local Similarity 21.2%; Pred. No. 26;
Matches 25; Conservative 14; Mismatches 38; Indels 41; Gaps 3;

QY 21 GVPQWG-----GGNNHNGGNSGPDSTLSI-----47
    |||||
Db 122 GVPYMGQGTTVTVSSGGGGGGGGSDIQLTQSHKFLSTSVGDRVSIICKASQDV 181
    |||||
QY 48 -----YQY--GSANALALQSDARKSETTITQSGYGNAGADYDQLVTRVYTHMAHAF 97
    |||||
Db 182 YNAVAVYQKPKGQSPKLLIYSASRYTGVPSPRFTGSGSGPDFTFTISSVQAEIDLAVYF 239
    |||||

Search completed: March 11, 2004, 19:14:20
Job time : 6.3 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:33:51 ; Search time 171.3 Seconds

(without alignments)
860.386 Million cell updates/sec

Title: US-09-543-407-30

Sequence: 1 MGLKLVAAALVSGSALA.....DSVMYRGFGNNATANDY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 6019581 seqs, 976053577 residues

Total number of hits satisfying chosen parameters: 6019581

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents AA Main:*
1: /cgn2_6/ptodata/2/paa/pctus_COMB.pep.*
2: /cgn2_6/ptodata/2/paa/US06_COMB.pep.*
3: /cgn2_6/ptodata/2/paa/US07_COMB.pep.*
4: /cgn2_6/ptodata/2/paa/US08_COMB.pep.*
5: /cgn2_6/ptodata/2/paa/US081_COMB.pep.*
6: /cgn2_6/ptodata/2/paa/US082_COMB.pep.*
7: /cgn2_6/ptodata/2/paa/US083_COMB.pep.*
8: /cgn2_6/ptodata/2/paa/US084_COMB.pep.*
9: /cgn2_6/ptodata/2/paa/US085_COMB.pep.*
10: /cgn2_6/ptodata/2/paa/US086_COMB.pep.*
11: /cgn2_6/ptodata/2/paa/US087_COMB.pep.*
12: /cgn2_6/ptodata/2/paa/US088_COMB.pep.*
13: /cgn2_6/ptodata/2/paa/US089_COMB.pep.*
14: /cgn2_6/ptodata/2/paa/US090_COMB.pep.*
15: /cgn2_6/ptodata/2/paa/US091_COMB.pep.*
16: /cgn2_6/ptodata/2/paa/US092_COMB.pep.*
17: /cgn2_6/ptodata/2/paa/US093_COMB.pep.*
18: /cgn2_6/ptodata/2/paa/US094_COMB.pep.*
19: /cgn2_6/ptodata/2/paa/US095_COMB.pep.*
20: /cgn2_6/ptodata/2/paa/US096_COMB.pep.*
21: /cgn2_6/ptodata/2/paa/US097A_COMB.pep.*
22: /cgn2_6/ptodata/2/paa/US097B_COMB.pep.*
23: /cgn2_6/ptodata/2/paa/US098_COMB.pep.*
24: /cgn2_6/ptodata/2/paa/US099_COMB.pep.*
25: /cgn2_6/ptodata/2/paa/US099B_COMB.pep.*
26: /cgn2_6/ptodata/2/paa/US100_COMB.pep.*
27: /cgn2_6/ptodata/2/paa/US101_COMB.pep.*
28: /cgn2_6/ptodata/2/paa/US102_COMB.pep.*
29: /cgn2_6/ptodata/2/paa/US103_COMB.pep.*
30: /cgn2_6/ptodata/2/paa/US104_COMB.pep.*
31: /cgn2_6/ptodata/2/paa/US106_COMB.pep.*
32: /cgn2_6/ptodata/2/paa/US107_COMB.pep.*
33: /cgn2_6/ptodata/2/paa/US60_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
------------	-------	-------------	--------	----	-------------

1	758	100.0	151	19	US-09-543-407-30	Sequence 30, Appl
2	700	91.1	151	19	US-09-543-407-24	Sequence 24, Appl
3	681	88.7	151	19	US-09-543-407-5	Sequence 5, Appl
4	676	88.0	151	6	US-08-233-642A-57	Sequence 57, Appl
5	659	85.8	151	19	US-09-543-407-14	Sequence 14, Appl
6	657	85.5	151	19	US-09-543-407-22	Sequence 22, Appl
7	613	79.8	151	19	US-09-543-407-28	Sequence 28, Appl
8	603	78.5	151	19	US-09-543-407-12	Sequence 12, Appl
9	600	78.1	151	19	US-09-543-407-26	Sequence 26, Appl
10	599	78.0	151	19	US-09-543-407-18	Sequence 18, Appl
11	597	77.7	151	19	US-09-543-407-20	Sequence 20, Appl
12	594	77.3	151	19	US-09-543-407-31	Sequence 31, Appl
13	566	73.7	151	19	US-09-543-407-16	Sequence 16, Appl
14	517	67.3	151	19	US-09-543-407-7	Sequence 7, Appl
15	514	66.9	151	13	US-08-978-878-4	Sequence 4, Appl
16	514	66.9	151	21	US-09-741-873B-4	Sequence 2, Appl
17	512	66.7	151	33	US-60-352-946-2	Sequence 2, Appl
18	512	66.7	151	33	US-60-444-371-2	Sequence 2, Appl
19	496	64.6	120	6	US-08-233-642A-55	Sequence 55, Appl
20	475	61.8	158	16	US-09-252-691-5834	Sequence 5834, Ap
21	475	61.8	158	16	US-09-252-691C-5834	Sequence 5834, Ap
22	475	61.8	158	30	US-10-417-886-5834	Sequence 34, Appl
23	462	60.2	109	19	US-09-543-407-34	Sequence 34, Appl
24	436	56.8	131	13	US-08-978-878-2	Sequence 2, Appl
25	436	56.8	131	21	US-09-741-873B-2	Sequence 2, Appl
26	332	43.2	109	19	US-09-543-407-35	Sequence 35, Appl
27	310	40.4	68	19	US-09-543-407-37	Sequence 37, Appl
28	237	30.9	48	19	US-09-543-407-39	Sequence 39, Appl
29	217.5	28.3	70	19	US-09-543-407-32	Sequence 32, Appl
30	114.5	14.9	186	16	US-09-252-691-5833	Sequence 5833, Ap
31	114.5	14.9	186	16	US-09-252-691C-5833	Sequence 5833, Ap
32	114.5	14.9	186	30	US-10-417-886-5833	Sequence 30, Appl
33	104.5	13.6	445	19	US-09-543-407-8	Sequence 8, Appl
34	103	13.4	445	29	US-10-369-493-20638	Sequence 20638, A
35	103	13.4	445	33	US-60-360-039-20638	Sequence 20638, A
36	101.5	13.0	520	1	PCT-US02-18262-21	Sequence 21, Appl
37	99.5	12.2	151	19	US-09-543-407-6	Sequence 6, Appl
38	98	12.8	145	21	US-09-739-449-8854	Sequence 8854, Ap
39	98	12.8	145	23	US-09-803-110-8854	Sequence 8854, Ap
40	98	12.8	375	21	US-09-733-089-22325	Sequence 22325, A
41	98	12.8	563	23	US-09-816-660-23325	Sequence 23325, A
42	98	12.8	563	21	US-09-733-089-23533	Sequence 23533, A
43	98	12.8	563	23	US-09-816-660-23533	Sequence 23533, A
44	96.5	12.6	313	21	US-09-708-427-9129	Sequence 9129, Ap
45	96.5	12.6	1249	30	US-10-455-719-358	Sequence 358, App

ALIGNMENTS

RESULT 1
US-09-543-407-30
Sequence 30, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collins, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIBRILLAR SYSTEM FOR PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543,407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 30
LENGTH: 151
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afgA
OTHER INFORMATION: sequence containing the replicon fragment
OTHER INFORMATION: encoding PT3 from GP63 of *Leishmania* major.

US-09-543-407-30

Query Match 100.0%; Score 768; DB 19; Length 151;
Best Local Similarity 100.0%; Pred. No. 7e-73;
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLKVAAFAAIVVSGSALAGVVPQWGGGNNHGGNNSGPDSTLSIYQYGSANAALALQ 60
DB 1 MKLKVAAFAAIVVSGSALAGVVPQWGGGNNHGGNNSGPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITGSGYNGADVGGADNSTIELTQNGFRNNATYDQVTRVTHMAHAG 120
DB 61 SDARKSETTITGSGYNGADVGGADNSTIELTQNGFRNNATYDQVTRVTHMAHAG 120
QY 121 NNALVNOTASDSSVWVROYGFGNNATANQY 151
DB 121 NNALVNOTASDSSVWVROYGFGNNATANQY 151

RESULT 2

US-09-543-407-24

; Sequence 24, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIBRILLAR SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543.407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE INFORMATION: Recombinant *Salmonella enteritidis* 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding P73 from GP63 of *Leishmania major*.
US-09-543-407-24

Query Match 91.1%; Score 700; DB 19; Length 151;
Best Local Similarity 89.9%; Pred. No. 1.2e-65;
Matches 143; Conservative 0; Mismatches 0; Indels 16; Gaps 2;

QY 1 MKLKVAAFAAIVVSGSALAGVVPQWGGGNNHGGNNSGPDSTLSIYQYGSANAALALQ 60
DB 1 MKLKVAAFAAIVVSGSALAGVVPQWGGGNNHGGNNSGPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITGSGYNGADVGGADNSTIELTQNGFRNNATYDQVTRVTHMAHAG 112
DB 61 SDARKSETTITGSGYNGADVGGADNSTIELTQNGFRNNATYDQVTRVTHMAHAG 112
QY 113 HEMAHAGNNALVNOTASDSSVWVROYGFGNNATANQY 151
DB 121 HEMAHAGNNALVNOTASDSSVWVROYGFGNNATANQY 151

RESULT 3

US-09-543-407-5

; Sequence 5, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIBRILLAR SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543.407

; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 151
; TYPE: PRT
; ORGANISM: *Salmonella enteritidis*
US-09-543-407-5

Query Match 88.7%; Score 681; DB 19; Length 151;
Best Local Similarity 90.7%; Pred. No. 1.2e-63;
Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

QY 1 MKLKVAAFAAIVVSGSALAGVVPQWGGGNNHGGNNSGPDSTLSIYQYGSANAALALQ 60
DB 1 MKLKVAAFAAIVVSGSALAGVVPQWGGGNNHGGNNSGPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITGSGYNGADVGGADNSTIELTQNGFRNNATYDQVTRVTHMAHAG 120
DB 61 SDARKSETTITGSGYNGADVGGADNSTIELTQNGFRNNATYDQVTRVTHMAHAG 120
QY 121 NNALVNOTASDSSVWVROYGFGNNATANQY 151
DB 121 NNALVNOTASDSSVWVROYGFGNNATANQY 151

RESULT 4

US-08-233-642A-57

; Sequence 57, Application US/08233642A
; GENERAL INFORMATION:
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Cloutier, Sharon C.
; APPLICANT: Doran, James L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR *SALMONELLA*-
; NUMBER OF SEQUENCES: 58
; TITLE OF INVENTION: BASED VACCINES
; NUMBER OF SEQUENCES: -
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,642A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERY
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 151 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-233-642A-57

Query Match 88.0%; Score 676; DB 6; Length 151;
Best Local Similarity 90.1%; Pred. No. 4.2e-63;
Matches 136; Conservative 2; Mismatches 13; Indels 0; Gaps 0;

```

Qy      1 MLLKVAAPAAIVVSGSALAGVVPQWGGGNNHGGSSGPDSTLSIYQGSANALALQ 60
        |||
Db      1 MLLKVAAPAAIVVSGSALAGVVPQWGGGNNHGGSSGPDSTLSIYQGSANALALQ 60
Qy      61 SPARKSETTITGSGYNGADVGGADNSTIELTONGFRNNATYDQLVTRVTHMAHAAG 120
        |||
Db      61 SPARKSETTITGSGYNGADVGGADNSTIELTONGFRNNATYDQLVTRVTHMAHAAG 120
Qy      121 NNAALVNOTASDSSVMVROVGFNNATANOY 151
        |||
Db      121 NNPALVNOTASDSSVMVROVGFNNATANOY 151

```

```

RESULT 5
US-09-543-407-14
; Sequence 14, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FaastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GPe3 of Leishmania major.
US-09-543-407-14

```

```

Query Match      85.8%; Score 659; DB 19; Length 151;
Best Local Similarity 81.9%; Pred. No. 2.7e-61;
Matches 136; Conservative 0; Mismatches 0; Indels 30; Gaps 2;
Qy      1 MLLKVAAPAAIVVSGSALAGVVPQWGGGNNHGGSSGPDSTLSIYQGSANALALQ 60
        |||
Db      1 MLLKVAAPAAIVVSGSALAGVVPQWGGGNNHGGSSGPDSTLSIYQGSANALALQ 60
Qy      61 SPARKSETTITGSGYNGADVGGADNSTIELTONGFRNNATYDQLVTRVTHMAHAAG 120
        |||
Db      61 SPARKSETTITGSGYNGADVGGADNSTIELTONGFRNNATYDQLVTRVTHMAHAAG 120
Qy      106 LVTRVTHMAHAAGNNALVNOTASDSSVMVROVGFNNATANOY 151
        |||
Db      121 LVTRVTHMAHAAGNNALVNOTASDSSVMVROVGFNNATANOY 151

```

```

RESULT 6
US-09-543-407-22
; Sequence 22, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FaastSeq for Windows Version 4.0
; SEQ ID NO 22
; LENGTH: 151

```

```

; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GPe3 of Leishmania major.
US-09-543-407-22

```

```

Query Match      85.5%; Score 657; DB 19; Length 151;
Best Local Similarity 81.9%; Pred. No. 4.3e-61;
Matches 136; Conservative 0; Mismatches 0; Indels 30; Gaps 2;
Qy      1 MLLKVAAPAAIVVSGSALAGVVPQWGGGNNHGGSSGPDSTLSIYQGSANALALQ 60
        |||
Db      1 MLLKVAAPAAIVVSGSALAGVVPQWGGGNNHGGSSGPDSTLSIYQGSANALALQ 60
Qy      61 SPARKSETTITGSGYNGADVGGADNSTIELTONGFRNNATYDQLVTRVTHMAHAAG 118
        |||
Db      61 SPARKSETTITGSGYNGADVGGADNSTIELTONGFRNNATYDQLVTRVTHMAHAAG 105

```

```

Qy      119 -----GGNNALVNOTASDSSVMVROVGFNNATANOY 151
        |||
Db      106 NNAKSDITVGGYGGNNALVNOTASDSSVMVROVGFNNATANOY 151

```

```

RESULT 7
US-09-543-407-28
; Sequence 28, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FaastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GPe3 of Leishmania major.
US-09-543-407-28

```

```

Query Match      79.8%; Score 613; DB 19; Length 151;
Best Local Similarity 74.6%; Pred. No. 2e-56;
Matches 129; Conservative 0; Mismatches 0; Indels 44; Gaps 2;
Qy      1 MLLKVAAPAAIVVSGSALAGVVPQWGGGNNHGGSSGPDSTLSIYQGSANALALQ 60
        |||
Db      1 MLLKVAAPAAIVVSGSALAGVVPQWGGGNNHGGSSGPDSTLSIYQGSANALALQ 60
Qy      61 SPARKSETTITGSGYNGADVGGADNSTIELTONGFRNNATYDQLVTRVTHMAHAAG 118
        |||
Db      61 SPARKSETTITGSGYNGADVGGADNSTIELTONGFRNNATYDQLVTRVTHMAHAAG 98

```

```

Qy      119 -----GGNNALVNOTASDSSVMVROVGFNNATANOY 151
        |||
Db      99 NNAKSDITVGGYGGNNALVNOTASDSSVMVROVGFNNATANOY 151

```

```

RESULT 8
US-09-543-407-12
; Sequence 12, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.

```

```

; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-12

```

```

Query Match      78.5%; Score 603; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 2.4e-55;
Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;

```

```

QY 1 MKLKVAAFAAIVVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSTIYQGSANAALALQ 60
DB 1 MKLKVAAFAAIVVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSTIYQGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGGAGDNSTIELTONGFRNNATYDQLVTRVVTHEMAHAGG 120
DB 61 SDARKSETTITQSGYNGADVGGAGDNSTIELTONGFRNNATYDQLVTRVVTHEMAHAGG 120
QY 121 NNAALVNOTASDSSVMRVQVFGNNATANQY 151
DB 121 NNAALVNOTASDSSVMRVQVFGNNATANQY 151

```

```

RESULT 9
US-09-543-407-26
; Sequence 26, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-26

```

```

Query Match      78.1%; Score 600; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 4.9e-55;
Matches 123; Conservative 4; Mismatches 24; Indels 0; Gaps 0;

```

```

QY 1 MKLKVAAFAAIVVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSTIYQGSANAALALQ 60
DB 1 MKLKVAAFAAIVVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSTIYQGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGGAGDNSTIELTONGFRNNATYDQLVTRVVTHEMAHAGG 120
DB 61 SDARKSETTITQSGYNGADVGGAGDNSTIELTONGFRNNATYDQLVTRVVTHEMAHAGG 120

```

```

QY 121 NNAALVNOTASDSSVMRVQVFGNNATANQY 151
DB 121 NNAALVNOTASDSSVMRVQVFGNNATANQY 151

```

```

RESULT 10
US-09-543-407-18
; Sequence 18, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-18

```

```

Query Match      78.0%; Score 599; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 6.3e-55;
Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;

```

```

QY 1 MKLKVAAFAAIVVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSTIYQGSANAALALQ 60
DB 1 MKLKVAAFAAIVVSGSALAGVVPQWGGGNNHNGGNSGPDSTLSTIYQGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGGAGDNSTIELTONGFRNNATYDQLVTRVVTHEMAHAGG 120
DB 61 SDARKSETTITQSGYNGADVGGAGDNSTIELTONGFRNNATYDQLVTRVVTHEMAHAGG 120
QY 121 NNAALVNOTASDSSVMRVQVFGNNATANQY 151
DB 121 NNAALVNOTASDSSVMRVQVFGNNATANQY 151

```

```

RESULT 11
US-09-543-407-20
; Sequence 20, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-20

```

```

Query Match      77.7%; Score 597; DB 19; Length 151;

```


Best Local Similarity 81.5%; Pred. No. 1e-54;
Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLKVAAFAAIVVSGSALAGVPQWGGGNNHGGSSGPDSTLSIYOGSANAALALQ 60
DB 1 MKLKVAAFAAIVVSGSALAGVPQWGGGNNHGGSSGPDSTLSIYOGSANAALALQ 60

QY 61 SDARKSETTITGSGYNGADVGGADNSTIELTONGFRNNATYDQVTRVVTHEMAHAGG 120
DB 61 SDARKSETTITGSGYNGADVGGADNSTIELTONGFRNNATYDQVTRVVTHEMAHAGG 120

QY 121 NNAALVNOTASDSSVWVROVGFNNATANOY 151
DB 121 NNAALVNOTASDSSVWVROVGFNNATANOY 151

RESULT 12
US-09-543-407-31
Sequence 31, Application US/09543407

GENERAL INFORMATION:
APPLICANT: White, Aaron P.

APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen

APPLICANT: Kay, William W.

TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
FILE REFERENCE: 920043.406

CURRENT APPLICATION NUMBER: US/09/543,407
CURRENT FILING DATE: 2000-04-05

NUMBER OF SEQ ID NOS: 59
SOFTWARE: FaastSeq for Windows Version 4.0

SEQ ID NO 31
LENGTH: 131

TYPE: PRT

ORGANISM: Salmonella enteritidis
US-09-543-407-31

Query Match
Best Local Similarity 77.3%; Score 594; DB 19; Length 131;
Matches 117; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

QY 21 GVPWQGGGNNHGGSSGPDSTLSIYOGSANAALALQSDARKSETTITGSGYNGAD 80
DB 1 GVPWQGGGNNHGGSSGPDSTLSIYOGSANAALALQSDARKSETTITGSGYNGAD 60

QY 81 VGGADNSTIELTONGFRNNATYDQVTRVVTHEMAHAGGNNALVNOTASDSSVWVROV 140
DB 61 VGGADNSTIELTONGFRNNATYDQVTRVVTHEMAHAGGNNALVNOTASDSSVWVROV 120

QY 141 GFGNNATANOY 151
DB 121 GFGNNATANOY 131

RESULT 13
US-09-543-407-16
Sequence 16, Application US/09543407

GENERAL INFORMATION:
APPLICANT: White, Aaron P.

APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen

APPLICANT: Kay, William W.

TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
FILE REFERENCE: 920043.406

CURRENT APPLICATION NUMBER: US/09/543,407
CURRENT FILING DATE: 2000-04-05

NUMBER OF SEQ ID NOS: 59
SOFTWARE: FaastSeq for Windows Version 4.0

SEQ ID NO 16
LENGTH: 151

TYPE: PRT
ORGANISM: Artificial Sequence

FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-16

Query Match
Best Local Similarity 73.7%; Score 566; DB 19; Length 151;
Matches 122; Conservative 5; Mismatches 24; Indels 0; Gaps 0;

QY 1 MKLKVAAFAAIVVSGSALAGVPQWGGGNNHGGSSGPDSTLSIYOGSANAALALQ 60
DB 1 MKLKVAAFAAIVVSGSALAGVPQWGGGNNHGGSSGPDSTLSIYOGSANAALALQ 60

QY 61 SDARKSETTITGSGYNGADVGGADNSTIELTONGFRNNATYDQVTRVVTHEMAHAGG 120
DB 61 SDARKSETTITGSGYNGADVGGADNSTIELTONGFRNNATYDQVTRVVTHEMAHAGG 120

QY 121 NNAALVNOTASDSSVWVROVGFNNATANOY 151
DB 121 NNAALVNOTASDSSVWVROVGFNNATANOY 151

RESULT 14
US-09-543-407-7
Sequence 7, Application US/09543407

GENERAL INFORMATION:
APPLICANT: White, Aaron P.

APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen

APPLICANT: Kay, William W.

TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
FILE REFERENCE: 920043.406

CURRENT APPLICATION NUMBER: US/09/543,407
CURRENT FILING DATE: 2000-04-05

NUMBER OF SEQ ID NOS: 59
SOFTWARE: FaastSeq for Windows Version 4.0

SEQ ID NO 7
LENGTH: 151

TYPE: PRT

ORGANISM: Escherichia coli
US-09-543-407-7

Query Match
Best Local Similarity 67.3%; Score 517; DB 19; Length 151;
Matches 105; Conservative 17; Mismatches 29; Indels 0; Gaps 0;

QY 1 MKLKVAAFAAIVVSGSALAGVPQWGGGNNHGGSSGPDSTLSIYOGSANAALALQ 60
DB 1 MKLKVAAFAAIVVSGSALAGVPQWGGGNNHGGSSGPDSTLSIYOGSANAALALQ 60

QY 61 SDARKSETTITGSGYNGADVGGADNSTIELTONGFRNNATYDQVTRVVTHEMAHAGG 120
DB 61 SDARKSETTITGSGYNGADVGGADNSTIELTONGFRNNATYDQVTRVVTHEMAHAGG 120

QY 121 NNAALVNOTASDSSVWVROVGFNNATANOY 151
DB 121 NNAALVNOTASDSSVWVROVGFNNATANOY 151

RESULT 15
US-08-978-878-4
Sequence 4, Application US/08978878

GENERAL INFORMATION:
APPLICANT: NORMARK, Stellan

APPLICANT: OLSEN, Arne

TITLE OF INVENTION: FIBRONECTIN BINDING PROTEIN AS WELL AS ITS PREPARATION
FILE REFERENCE: 012889-081

CURRENT APPLICATION NUMBER: US/08/978,878
CURRENT FILING DATE: 1997-11-26

EARLIER APPLICATION NUMBER: SE 8801723-1
EARLIER FILING DATE: 1988-05-06

; EARLIER APPLICATION NUMBER: US 07/347,189
; EARLIER FILING DATE: 1989-05-04
; EARLIER APPLICATION NUMBER: US 07/789,437
; EARLIER FILING DATE: 1991-11-06
; EARLIER APPLICATION NUMBER: US 07/970,846
; EARLIER FILING DATE: 1992-11-03
; EARLIER APPLICATION NUMBER: US 08/187,865
; EARLIER FILING DATE: 1994-01-28
; EARLIER APPLICATION NUMBER: US 08/318,519
; EARLIER FILING DATE: 1994-10-05
; EARLIER APPLICATION NUMBER: US 08/495,959
; EARLIER FILING DATE: 1995-06-28
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-08-978-878-4

Query Match 66.9%; Score 514; DB 13; Length 151;
Best Local Similarity 68.9%; Pred. No. 6.7e-46;
Matches 104; Conservative 18; Mismatches 29; Indels 0; Gaps 0;

QY 1 MKLIKVAFAAIVVSGSALAGVVPQWGGGNNHGGNSGSPDSTLSIYOGSANAALALQ 60
DB 1 MKLIKVAFAAIVVSGSALAGVVPQYGGGNNHGGNSGPNSELNIYOGGNSALALQ 60
QY 61 SDARKSETTTTQSGYGAGADYGGAGDNSTIELTONGFRNNATYDQVTRVVTHEMAHAG 120
DB 61 TDARNSDLITTOHGGGAGADYGGGSDSDSIDLTORGFGNSATLDQWNGKSEMTVKQFGG 120
QY 121 NNALVNOTASDSSVMTRVGVFGNNATANOY 151
DB 121 GNGAAVDQTAASSSVNVTQVGFNNATRAHOT 151

Search completed: March 11, 2004, 19:13:14
Job time : 171.3 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 18:35:05 ; Search time 5.3 Seconds
(without alignments)
376.014 Million cell updates/sec

Title: US-09-543-407-30

Perfect score: 768

Sequence: 1 MLLKVAAPAAIVSGSALA.....DSVMYROYFGNNATANOY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 125546 seqs, 13197846 residues

Total number of hits satisfying chosen parameters: 125546

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents AA New.*
1: /cgn2_6/prodata/1/paa/PCT_NEW_COMB.pep.*
2: /cgn2_6/prodata/1/paa/US06_NEW_COMB.pep.*
3: /cgn2_6/prodata/1/paa/US07_NEW_COMB.pep.*
4: /cgn2_6/prodata/1/paa/US08_NEW_COMB.pep.*
5: /cgn2_6/prodata/1/paa/US09_NEW_COMB.pep.*
6: /cgn2_6/prodata/1/paa/US10_NEW_COMB.pep.*
7: /cgn2_6/prodata/1/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	514	66.9	151	US-09-741-873C-4	Sequence 4, Appl
2	436	56.8	131	US-09-741-873C-2	Sequence 2, Appl
3	91.5	11.9	1327	PCT-US04-02338-49	Sequence 49, Appl
4	85	11.1	321	PCT-US04-05654-590	Sequence 590, Appl
5	82	10.7	443	US-10-100-683-7608	Sequence 7608, Ap
6	79	10.3	351	PCT-US04-05654-1755	Sequence 1755, Ap
7	77.5	10.1	1160	US-10-603-150-2	Sequence 2, Appl
8	76.5	10.0	386	US-10-100-683-10326	Sequence 10326, A
9	76.5	10.0	386	US-10-100-683-10327	Sequence 10327, A
10	75.5	9.9	250	US-10-779-461-40	Sequence 40, Appl
11	75.5	9.8	382	US-10-771-241-299	Sequence 299, App
12	75.5	9.8	533	US-10-045-674A-551	Sequence 551, App
13	75	9.8	424	US-10-045-674A-551	Sequence 551, App
14	74	9.6	234	US-10-767-701-42696	Sequence 42696, A
15	73.5	9.6	624	PCT-US04-05654-1466	Sequence 1466, Ap
16	73	9.5	246	US-10-779-461-41	Sequence 41, Appl
17	72	9.4	434	US-10-045-674A-594	Sequence 594, App
18	71.5	9.3	175	US-10-767-701-35046	Sequence 35046, A
19	71.5	9.3	175	US-10-767-701-42696	Sequence 42696, A
20	71.5	9.3	386	US-10-767-701-42417	Sequence 42417, A
21	70.5	9.2	251	US-10-417-884A-5724	Sequence 5724, Ap
22	70.5	9.2	251	US-10-767-701-32417	Sequence 32417, A
23	70.5	9.2	273	US-10-627-556-392	Sequence 392, App
24	70.5	9.2	273	US-10-627-556-396	Sequence 396, App
25	70.5	9.2	505	US-10-627-556-400	Sequence 400, App
26	70.5	9.2	505	US-10-627-556-402	Sequence 402, App

27	70.5	9.2	521	US-10-627-556-398	Sequence 398, App
28	70.5	9.2	1048	PCT-US04-05654-2086	Sequence 2086, Ap
29	70	9.1	199	US-10-767-701-51583	Sequence 51583, A
30	70	9.1	220	US-10-767-701-52948	Sequence 52948, A
31	70	9.1	245	US-10-779-461-59	Sequence 59, Appl
32	69.5	9.0	175	US-10-767-701-59665	Sequence 59665, A
33	69.5	9.0	386	PCT-US04-05654-2526	Sequence 2526, Ap
34	69	9.0	205	US-10-767-701-40104	Sequence 40104, A
35	69	9.0	250	US-10-779-461-6	Sequence 6, Appl
36	69	9.0	251	US-10-779-461-30	Sequence 30, Appl
37	69	9.0	588	PCT-US04-05654-336	Sequence 336, App
38	69	9.0	588	US-10-764-425-187	Sequence 187, App
39	69	9.0	588	US-10-767-701-46258	Sequence 46258, A
40	69	9.0	620	US-10-771-241-364	Sequence 364, App
41	69	9.0	878	US-10-190-902B-7	Sequence 7, Appl
42	68.5	8.9	173	US-10-767-701-32246	Sequence 32246, A
43	68.5	8.9	191	US-10-767-701-47075	Sequence 47075, A
44	68.5	8.9	245	US-10-779-461-19	Sequence 19, Appl
45	68.5	8.9	245	US-10-779-461-19	Sequence 19, Appl

ALIGNMENTS

RESULT 1
US-09-741-873C-4
Sequence 4, Application US/09741873C
GENERAL INFORMATION:
APPLICANT: Normark, Stefan
APPLICANT: Olsen, Arne
TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
FILE REFERENCE: 012889-084
CURRENT APPLICATION NUMBER: US/09/741,873C
CURRENT FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: SE 8801723-1
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: US 08/978,878
PRIOR FILING DATE: 1997-11-26
PRIOR APPLICATION NUMBER: US 07/347,189
PRIOR FILING DATE: 1989-05-04
PRIOR APPLICATION NUMBER: US 07/789,437
PRIOR FILING DATE: 1991-11-06
PRIOR APPLICATION NUMBER: US 07/970,846
PRIOR FILING DATE: 1992-11-03
PRIOR APPLICATION NUMBER: US 08/187,865
PRIOR FILING DATE: 1994-01-28
PRIOR APPLICATION NUMBER: US 08/318,519
PRIOR FILING DATE: 1994-10-05
NUMBER OF SEQ ID NOS: 11
SOFTWARE: PatentIn version 3.0
SEQ ID NO 4
LENGTH: 151
TYPE: PRT
ORGANISM: Escherichia coli
US-09-741-873C-4

Query Match 66.9%; Score 514; DB 5; Length 151;
Best Local Similarity 68.9%; Pred. No. 2, le-38;
Matches 104; Conservative 18; Mismatches 29; Indels 0; Gaps 0;

QY	1	MLLKVAAPAAIVSGSALAGVVPQGGGNGGNSGSPDSTLTIYOGSANAALAIQ 60
DB	1	MLLKVAAPAAIVSGSALAGVVPQGGGNGGNSGSPDSTLTIYOGSANAALAIQ 60
QY	61	SNARSETTITGSGVNGADVQGGADNSTIETLQNGFRNATYDOLVTRVTHMAHAG 120
DB	61	TGARNSDLTITGSGVNGADVQGGADNSTIETLQNGFRNATYDOLVTRVTHMAHAG 120
QY	121	NNAALVQTAQSDSVYVROVFGNNATANOY 151
DB	121	GNGAIVDQTAQSDSVYVROVFGNNATANOY 151

```

RESULT 2
US-09-741-873C-2
; Sequence 2, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staefan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO: 2
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873C-2

Query Match      56.8%; Score 436; DB 5; Length 131;
Best Local Similarity 65.6%; Pred. No. 1,2e-31;
Matches 86; Conservative 17; Mismatches 28; Indels 0; Gaps 0;

QY 21 GVPWGGGNNHNGSGSPDSTLSTIYQGSANAAALALQSDARKSETTITQSGYNGAD 80
DB 1 GVPWGGGNNHNGSGSPDSTLSTIYQGSANAAALALQSDARKSETTITQSGYNGAD 60
QY 81 VGGAGDNSTIELTONGFRNNATYDQLVTRVYTHMAHAGNNALVNQTSASDSSVMROY 140
DB 61 VGGAGDNSTIELTONGFRNNATYDQLVTRVYTHMAHAGNNALVNQTSASDSSVMROY 120
QY 141 GFGNNATANOY 151
DB 121 GFGNNATANOY 131

RESULT 3
PCT-US04-02338-49
; Sequence 49, Application PC/TUS0402338
; GENERAL INFORMATION:
; APPLICANT: EXELIXIS, INC.
; TITLE OF INVENTION: MAPKs AS MODIFIERS OF THE APC AND AXIN PATHWAYS AND METHODS OF
; FILE REFERENCE: EX04-003C-PC
; CURRENT APPLICATION NUMBER: PCT/US04/02338
; CURRENT FILING DATE: 2004-01-28
; PRIOR APPLICATION NUMBER: US60/443,484
; PRIOR FILING DATE: 2003-01-29
; PRIOR APPLICATION NUMBER: US60/447,358
; PRIOR FILING DATE: 2003-02-11
; PRIOR APPLICATION NUMBER: US60/461,789
; PRIOR FILING DATE: 2003-04-10
; PRIOR APPLICATION NUMBER: US60/470,684
; PRIOR FILING DATE: 2003-05-14
; PRIOR APPLICATION NUMBER: US60/479,650
; PRIOR FILING DATE: 2003-06-19
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO: 49
; LENGTH: 1327

```

```

; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US04-02338-49

Query Match      11.9%; Score 91.5; DB 1; Length 1327;
Best Local Similarity 29.4%; Pred. No. 2.7;
Matches 42; Conservative 17; Mismatches 65; Indels 19; Gaps 6;

QY 6 VAAFAAI-VVSGALAGVDPQWGGNNHNGSGSPDSTLSTIYQGSANAAALALQSDAR 64
DB 99 VAAAPVPAVTSIAAGVAPNPAGSGSNSSPSSPTSS-SSSPSSPSSSLAESPEAA 157
QY 65 KSEITTT---QSGYNGADVGGADNSTIELTONGFRNNATYDQLVTRVYTHMAHAG 120
DB 158 GVSTAPLPGGAGPGNGVPAVSGALRELLACRND-----VSRV--KRLVDAA 206
QY 121 NNALVNQTSASDSSVMROYGFG 143
DB 207 VNAK--DMGRKSSPLHFAAGFG 227

RESULT 4
PCT-US04-05654-590
; Sequence 590, Application PC/TUS0405654
; GENERAL INFORMATION:
; APPLICANT: Sherman, Bradley K
; APPLICANT: Riechmann, Jose Luis
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Heard, Jacqueline E
; APPLICANT: Haake, Volker
; APPLICANT: Creelman, Robert A
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Adam, Luc J
; APPLICANT: Reuber, T. Lynne
; APPLICANT: Keddie, James
; APPLICANT: Dubeil, Ilt, Arnold N
; APPLICANT: Pineda, Omaira
; APPLICANT: Repetti, Peter
; APPLICANT: Century, Karen
; APPLICANT: Guttersen, Neal
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Broun, Pierre E
; APPLICANT: Kumimoto, Roderick W
; APPLICANT: Pilgrim, Marsha L
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES IN PLANTS
; FILE REFERENCE: MBI-0047 PCT
; CURRENT APPLICATION NUMBER: PCT/US04/05654
; CURRENT FILING DATE: 2004-03-03
; PRIOR APPLICATION NUMBER: 10/374,780
; PRIOR FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: 10/675,852
; PRIOR FILING DATE: 2003-09-30
; NUMBER OF SEQ ID NOS: 2950
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO: 590
; LENGTH: 321
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: G3384 Orthologous to G256
PCT-US04-05654-590

Query Match      11.1%; Score 85; DB 1; Length 321;
Best Local Similarity 26.3%; Pred. No. 1.8;
Matches 30; Conservative 15; Mismatches 41; Indels 28; Gaps 4;

QY 10 AATVSGALAGVDPW---GGGNNHNGSGSPDSTLSTIYQGSANAAALALQSDARKS 66
DB 199 AAVASADNIALILQGMWPGG---GGGNGKGPBAS-----GSTTATATQOOPQCS 248
QY 67 ETTITQSGYNGADVGGADNSTIELTONGFRNNATYDQLVTRVYTHMAHAG 120
DB 249 -----GGGASASASASGSAALAAATATQTPCSTSTSKATGCG 287

```



```

; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: US 60/056,664
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/043,314
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: US 60/047,632
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: US 60/056,892
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 13468
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10327
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-100-683-10327

Query Match          10.0%; Score 76.5; DB 6; Length 386;
Best Local Similarity 31.1%; Pred. No. 12;
Matches 28; Conservative 13; Mismatches 34; Indels 15; Gaps 5;

Cy      21 GVYPQ--WGGGNNHNGSGSPDSTLSIYQ-YGSANALALQSDARKSETTTTOSGY 75
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      185 GMPDQGAPWGCGG--NCGPPNFNTGTGAVAQPEYGSVRA-----SNNEGCTNPSPGS 237

Cy      76 GNGAD--VGGADNSTIELTQCNFRNAT 102
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      238 GGSSNSGGSGSGSOGSSGSSGSGNDNNNS 267

RESULT 10
US-10-779-461-40
; Sequence 40, Application US/10779461
; GENERAL INFORMATION:
; APPLICANT: Moricon, Philip A
; TITLE OF INVENTION: ANTIBODIES TO C-MET FOR THE TREATMENT OF CANCERS
; FILE REFERENCE: 00980/1
; CURRENT APPLICATION NUMBER: US/10/779,461
; CURRENT FILING DATE: 2004-02-13
; PRIOR APPLICATION NUMBER: 60/447,073
; PRIOR FILING DATE: 2003-02-13
; NUMBER OF SEQ ID NOS: 161
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 40
; LENGTH: 250
; TYPE: PRT
; ORGANISM: artificial
; FEATURE:
; OTHER INFORMATION: phage display generated human antibody
US-10-779-461-40

Query Match          9.9%; Score 76; DB 6; Length 250;
Best Local Similarity 30.6%; Pred. No. 7.9;
Matches 26; Conservative 8; Mismatches 31; Indels 20; Gaps 3;

Cy      9 FAIVVGSAALAGVPQMG-----GGNNHNGSGSPDSTLSIYQGSANALAL 59
       :|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      95 YCAISPRGLADVDFDWGCGTLVTVASGGGGSGSGSGG-----GSAQSALTQ 144

Cy      60 QSDARKSE-TTITOSGVGNQADVQ 83
       ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      145 PASASGPGQSTITICTGTSSDIKR 169

RESULT 11
US-10-771-241-299
; Sequence 299, Application US/10771241
; GENERAL INFORMATION:
; APPLICANT: Zykend, Judith
; APPLICANT: Forsyth, R. Aliyn
; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
; TITLE OF INVENTION: ESCHERICHTIA COLI
```


Query Match	9.6%;	Score 73.5;	DB 1;	Length 624
Best Local Similarity	26.1%;	Pred. No. 40;		

Search completed: March 11, 2004, 19:14:20
Job time : 5.3 secs